

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 2 年 9 月 1 9 日
Date of Application:

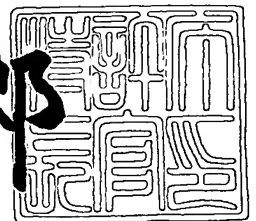
出 願 番 号 特 願 2 0 0 2 - 2 7 3 0 7 1
Application Number:
[ST. 10/C]: [J P 2 0 0 2 - 2 7 3 0 7 1]

出 願 人 セイコーエプソン株式会社
Applicant(s):

2 0 0 3 年 7 月 8 日

特許庁長官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特 2 0 0 3 - 3 0 5 4 2 5 1

【書類名】 特許願

【整理番号】 J0094471

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00

【発明者】

 【住所又は居所】 長野県諏訪市大和 3 丁目 3 番 5 号 セイコーエプソン株式会社内

 【氏名】 西谷 正信

【発明者】

 【住所又は居所】 長野県諏訪市大和 3 丁目 3 番 5 号 セイコーエプソン株式会社内

 【氏名】 宮澤 康永

【発明者】

 【住所又は居所】 長野県長野市栗田 1 5 0 - 8

 【氏名】 松本 弘

【発明者】

 【住所又は居所】 長野県長野市三輪 9 - 2 3 - 1 8 宇木住宅 1 - 1 0 4 号

 【氏名】 山本 一公

【特許出願人】

 【識別番号】 000002369

 【氏名又は名称】 セイコーエプソン株式会社

【代理人】

 【識別番号】 100095728

 【弁理士】

 【氏名又は名称】 上柳 雅誉

 【連絡先】 0 2 6 6 - 5 2 - 3 1 3 9

【選任した代理人】

【識別番号】 100107076

【弁理士】

【氏名又は名称】 藤網 英吉

【選任した代理人】

【識別番号】 100107261

【弁理士】

【氏名又は名称】 須澤 修

【手数料の表示】

【予納台帳番号】 013044

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0109826

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音響モデル作成方法および音声認識装置

【特許請求の範囲】

【請求項 1】 HMM（隠れマルコフモデル）を構成するそれぞれの状態のガウス分布数をそれぞれの状態ごとに最適化して、その最適化されたHMMを学習用音声データを用いて再学習してHMMを作成する音響モデル作成方法であって、

HMMを構成する複数の状態の各状態ごとに、ガウス分布数をある値から最大分布数までの複数種類の分布数に設定し、

この複数種類のガウス分布数に設定されたそれぞれの状態に対して、それぞれのガウス分布数ごとに記述長最小基準を用いて記述長を求め、

この記述長が最小となるガウス分布数を持つ状態をそれぞれの状態ごとに選択し、

このそれぞれの状態ごとに選択された記述長が最小となるガウス分布数を持つ状態によってそのHMMを構築し、その構築されたHMMを学習用音声データを用いて再学習する、

ことを特徴とする音響モデル作成方法。

【請求項 2】 前記記述長最小基準は、モデル集合 $\{1, \dots, i, \dots, I\}$ とデータ $\chi^N = \{\chi_1, \dots, \chi_N\}$ （ただし、 N はデータ長）が与えられたときのモデル i を用いた記述長 $l_i(\chi^N)$ が、一般的な式として、

【数 1】

$$l_i(\chi^N) = -\log P_{\theta^{(i)}}(x^N) + \frac{\beta_i}{2} \log N + \log I \quad (1)$$

$$\left(\begin{array}{l} \hat{\theta}^{(i)} : \text{モデル } i \text{ のパラメータ } \theta^{(i)} = \theta_1^{(i)}, \dots, \theta_{\beta_i}^{(i)} \text{ の最尤推定量} \\ \beta_i : \text{モデル } i \text{ の次元(自由度)} \end{array} \right)$$

で表され、この記述長を求める一般的な式において、前記モデル集合 $\{1, \dots, i, \dots, I\}$ は、ある HMM におけるある状態のガウス分布数がある値から最大分布数までの複数種類に設定された状態の集合であるとして考え、ここで、前記ガウス分布数の種類の数が I 種類 (I は $I \geq 2$ の整数) であるとき、前記 $1, \dots, i, \dots, I$ は、1 番目の種類から I 番目の種類までのそれぞれの種類を特定するための符号であって、前記 (1) 式を、前記 $1, \dots, i, \dots, I$ のうちの i 番目の分布数の種類を持つ状態の記述長を求める式として用いることを特徴とする請求 1 記載の音響モデル作成方法。

【請求項 3】 前記記述長を求める一般的な式において、右辺の第 2 項に重み係数 α を乗じることを特徴とする請求項 2 記載の音響モデル作成方法。

【請求項 4】 前記記述長を求める一般的な式において、右辺の第 2 項に重み係数 α を乗じ、かつ、右辺の第 3 項を省略したことを特徴とする請求項 2 記載の音響モデル作成方法。

【請求項 5】 前記データ χ^N は、前記ある値から最大分布数までのうちの任意のガウス分布数をそれぞれの状態に持つ HMM を用い、その HMM のそれぞれの状態と多数の学習用音声データとをそれぞれの状態ごとに時系列的な対応付けを行って得られるそれぞれの学習用音声データの集合であることを特徴とする請求項 2 から 4 のいずれかに記載の音響モデル作成方法。

【請求項 6】 前記任意のガウス分布数は、前記最大分布数であることを特徴とする請求項 5 記載の音響モデル作成方法。

【請求項 7】 前記 HMM は、音節 HMM であることを特徴とする請求項 1 から 6 のいずれかに記載の音響モデル作成方法。

【請求項 8】 前記音節 HMM において、同一子音や同一母音を持つ複数の音節 HMM に対し、これらの音節 HMM を構成する状態のうち、同一子音を有する音節 HMM 同士においては、それら音節 HMM における初期状態またはこの初期状態を含む少なくとも 2 つの状態を共有し、同一母音を有する音節 HMM 同士においては、それら音節 HMM における自己ループを有する状態の最終状態またはこの最終状態を含む少なくとも 2 つの状態を共有することを特徴とする請求項 7 記載の音響モデル作成方法。

【請求項9】 入力音声の特徴分析して得られた特徴データに対し音響モデルとしてHMM（隠れマルコフモデル）を用いて前記入力音声を認識する音声認識装置であって、

前記音響モデルとしてのHMMとして、前記請求項1から請求項8のいずれかに記載の音響モデル作成方法によって作成されたHMMを用いることを特徴とする音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音響モデルとして混合連続分布型HMM（隠れマルコフモデル）を作成する音響モデル作成方法およびこの音響モデルを用いた音声認識装置に関する。

【0002】

【従来の技術】

音声認識においては、音響モデルとして音素HMMや音節HMMを用い、この音素HMMや音節HMMを連結して、単語や文節、文といった単位の音声言語を認識する方法が一般的に行われている。特に最近、より高い認識性能を持つ音響モデルとして、混合連続分布型HMMが広く使われている。

【0003】

一般的に、HMMは1個から10個の状態とその間の状態遷移から構成されている。各状態でのシンボル（ある時刻の音声特徴ベクトル）の出現確率の計算において、混合連続分布型HMMでは、ガウス分布数が多いほど認識精度が高くなるが、ガウス分布数が多ければその分、パラメータ数も多くなり計算量やメモリ使用量が増大するという問題がある。これは処理能力の低いプロセッサや小容量のメモリを用いざるを得ない安価な機器に音声認識機能を搭載する場合、特に大きな問題となる。

【0004】

また、一般的な混合連続分布型HMMでは、すべての音素（または音節）HMMの全状態でガウス分布数が同じであるため、学習用音声データが少ない音素（

または音節) HMMでは過学習が起こり、該当する音素(音節)で認識性能が低くなるという問題もある。

【0005】

このように、混合連続分布型HMMではそれぞれの音素(または音節)の全状態においてガウス分布数が一定であるのが一般的であり、認識精度を高めるため、それぞれの状態におけるガウス分布数はある程度の数が必要である。しかしながら、上述したように、ガウス分布数が多ければその分、パラメータ数も多くなり計算量やメモリ使用量が増大するという問題もあるので、むやみにガウス分布数を増やすことはできないのが現状である。

【0006】

そこで、音素(または音節) HMMにおいて、それぞれの状態ごとにガウス分布数を異ならせる、つまり、それぞれの状態ごとにガウス分布数を最適化することが考えられる。たとえば、音節HMMを例にとれば、ある音節HMMを構成する各状態において、認識に大きく影響を与える部分の状態とそれほど大きな影響を与えない状態が存在することを考慮して、認識に大きく影響を与える部分の状態はガウス分布数を多くし、認識にそれほど大きな影響を与えない状態はガウス分布数を少なくすることが考えられる。

【0007】

このように、音素(または音節) HMMにおいてそれぞれの状態ごとにガウス分布数を最適化しようとする技術の一例として、「“MDL基準を用いたHMMサイズの削減” 篠田浩一、磯健一、2002年春季研究発表会 日本音響学会講演論文集 2002年3月、79～80頁」がある。

【0008】

【非特許文献1】

“MDL基準を用いたHMMサイズの削減” 篠田浩一、磯健一、2002年春季研究発表会 日本音響学会講演論文集 2002年3月、79～80頁

【0009】

【発明が解決しようとする課題】

この従来技術は、各状態において、認識に対する寄与の少ない部分におけるガウス分布数を削減することについて記載されており、簡単に言えば、十分な学習用音声データ量で学習された大きなガウス分布数を持つHMMを用意し、その状態ごとのガウス分布数の木構造を作成し、各状態ごとに記述長最小（MDL: Minimum Description Length）基準を最小にするガウス分布数の集合を選ぶものである。

【0010】

この従来技術によれば、確かに、音素（または音節）HMMにおいてそれぞれの状態ごとにガウス分布数を効果的に削減することができ、しかも、それぞれの状態におけるガウス分布数の最適化が可能となり、ガウス分布数の削減によるパラメータ数の削減を可能としながらも高い認識率を維持できると考えられる。

【0011】

しかしながら、この従来技術は、状態ごとのガウス分布数数の木構造を作成し、その木構造の分布の中からMDL基準を最小とするガウス分布集合（ノードの組み合わせ）を選択するというものであるため、ある状態において最適なガウス分布数を得るためのノードの組み合わせ数は極めて多く、それぞれの組み合わせごとに記述長を求めるために多くの演算を行う必要がある。

【0012】

なお、このMDL基準は、モデル集合 $\{1, \dots, i, \dots, I\}$ とデータ $\chi^N = \{\chi_1, \dots, \chi_N\}$ が与えられたときのモデル i を用いた記述長 $l_i(\chi^N)$ が、特許請求の範囲に記載した（1）式のように定義される。

【0013】

MDL基準は、この記述長 $l_i(\chi^N)$ が最小であるモデルが最適なモデルであるとしているが、この従来技術では、ノードの組み合わせが極めて多くなる可能性があることから、最適なガウス分布集合を選択する際に、その（1）式を近似した記述長計算式を用いて、ノードの組み合わせで構成されるガウス分布集合の記述長を求めている。このように、近似式によってノードの組み合わせで構成されるガウス分布集合の記述長が求められると、求められた結果の精度に多少の問題が生じる場合もあると考えられる。

【0014】

本発明は、それぞれの音素（または音節）HMMの各状態ごとのガウス分布数をMDL基準を用い、少ない演算量で精度よく最適な分布数の設定を可能とすることで、少ない演算量で高い認識性能が得られるHMMの作成が可能な音響モデル作成方法を提供するとともに、その音響モデルを用いることにより、演算能力やメモリ容量などハードウェア資源に大きな制約のある安価なシステムに適用できる音声認識装置を提供することを目的としている。

【0015】

【課題を解決するための手段】

上述した目的を達成するために、本発明の音響モデル作成方法は、HMMを構成するそれぞれの状態のガウス分布数をそれぞれの状態ごとに最適化して、その最適化されたHMMを学習用音声データを用いて再学習してHMMを作成する音響モデル作成方法であって、HMMを構成する複数の状態の各状態ごとに、ガウス分布数がある値から最大分布数までの複数種類のガウス分布数に設定し、この複数種類のガウス分布数に設定されたそれぞれの状態に対して、それぞれのガウス分布数ごとに記述長最小基準を用いて記述長を求め、この記述長が最小となるガウス分布数を持つ状態をそれぞれの状態ごとに選択し、このそれぞれの状態ごとに選択された記述長が最小となるガウス分布数を持つ状態によってそのHMMを構築し、その構築されたHMMを学習用音声データを用いて再学習するようにしている。

【0016】

このような音響モデル作成方法において、前記記述長最小基準は、モデル集合 $\{1, \dots, i, \dots, I\}$ とデータ $\chi^N = \{\chi_1, \dots, \chi_N\}$ （ただし、 N はデータ長）が与えられたときのモデル i を用いた記述長 $l_i(\chi^N)$ が、一般的な式として前記（1）式で表され、この記述長を求める一般的な式において、前記モデル集合 $\{1, \dots, i, \dots, I\}$ は、あるHMMにおけるある状態のガウス分布数がある値から最大分布数までの複数種類に設定された状態の集合であるとして考え、ここで、前記ガウス分布数の種類の数が I 種類（ I は $I \geq 2$ の整数）であるとき、前記 $1, \dots, i, \dots, I$ は、1番目の種類

から I 番目の種類までのそれぞれの種類を特定するための符号であって、前記 (1) 式を、前記 1, \dots , i, \dots , I のうちの i 番目の分布数の種類を持つ状態の記述長を求める式として用いるようにしている。

【0017】

また、前記記述長を求める一般的な式において、右辺の第 2 項に重み係数 α を乗じるようにしている。

【0018】

また、前記記述長を求める一般的な式において、右辺の第 2 項に重み係数 α を乗じ、かつ、右辺の第 3 項を省略するようにしてもよい。

【0019】

また、前記データ χ^N は、前記ある値から最大分布数までのうちのある任意のガウス分布数をそれぞれの状態に持つ HMM を用い、その HMM のそれぞれの状態と多数の学習用音声データとをそれぞれの状態ごとに時系列的な対応付けを行って得られるそれぞれの学習用音声データの集合であるとしている。なお、このとき、前記任意のガウス分布数は、前記最大分布数とすることが好ましい。

【0020】

また、前記 HMM が音節 HMM である場合、同一子音や同一母音を持つ複数の音節 HMM に対し、これらの音節 HMM を構成する状態のうち、同一子音を有する音節 HMM 同士においては、それら音節 HMM における初期状態またはこの初期状態を含む少なくとも 2 つの状態を共有し、同一母音を有する音節 HMM 同士においては、それら音節 HMM における自己ループを有する状態の最終状態またはこの最終状態を含む少なくとも 2 つの状態を共有することもできる。

【0021】

また、本発明の音声認識装置は、入力音声の特徴分析して得られた特徴データに対し音響モデルとして HMM を用いて前記入力音声を認識する音声認識装置であって、前記音響モデルとしての HMM として、上述の音響モデル作成方法によって作成された HMM を用いるようにしている。

【0022】

このように本発明では、それぞれの状態ごとにガウス分布数（以下では、単に

分布数という)の最適化を行うために、HMMを構成する複数の状態ごとに、ガウス分布数がある値から最大分布数まで複数種類の分布数に設定し、このガウス分布数がある値から最大分布数まで設定された状態に対して、分布数がある値から最大分布数のどの分布数が最適であるかを記述長最小基準を用いて選択し、記述長が最小となる分布数を持つ状態によってそれぞれのHMMを構築し、その構築されたそれぞれのHMMに対して学習用音声データを用いて再学習するようにしている。これによって、少ない演算量で最適な分布数の設定が可能となり、少ない演算量で高い認識性能が得られるHMMを作成することができる。

【0023】

特に、本発明の場合、分布数がある値から最大分布数までの中から最適な分布数を持つ状態を選択するというものであるため、たとえば、ある状態における分布数の種類を7種類とすれば、1つの状態において記述長を求める計算を7回行って、その中から記述長最小となる状態を選択すればよいので、少ない演算量で最適な分布数設定が可能となることが特徴の1つである。

【0024】

また、本発明ではMDL基準におけるモデル集合 $\{1, \dots, i, \dots, I\}$ は、あるHMMにおけるある状態のガウス分布数がある値から最大分布数までの複数種類に設定された状態の集合であるとして考え、前述の(1)式を、 $1, \dots, i, \dots, I$ のうちの*i*番目の分布数の種類を持つ状態の記述長を求める式として用いるようにしているので、ある状態における分布数がある値から最大分布数までの様々な分布数の種類に設定したとき、それぞれの分布数に設定された状態の記述長を容易に計算することができる。そして、その結果から、記述長最小となる分布数を求めることで、その状態における最適な分布数を設定することができる。

【0025】

また、記述長を求める一般的な式において、右辺の第2項に重み係数 α を乗じるようにしている。これによって、重み係数 α を可変することによって、第2項の単調増加の傾きを可変(α を大きくするほど傾きが大きくなる)することができる、記述長 $l_i(\chi^N)$ を可変させることができるので、たとえば、 α を大きく

すると、分布数がより小さい場合に記述長 $l_i(\chi^N)$ が最小になるように調整することができる。

【0026】

また、記述長を求める一般的な式において、右辺の第2項に重み係数 α を乗じ、かつ、定数を表す右辺の第3項を省略することによって、記述長を求める計算をより簡略化することができる。

【0027】

また、ある任意の分布数をそれぞれの状態に持つHMMを用い、そのHMMのある状態とそのHMMに対応する多数の学習用音声データとを時系列的な対応付け（たとえばビタビアライメント）を行い、その対応付けられた区間に対応するそれぞれの学習用音声データの集合を（1）式のデータ χ^N として用いている。このように、ある任意の分布数をそれぞれの状態に持つHMMを用い、そのHMMのある状態とそのHMMに対応する多数の学習用音声データとを時系列的な対応付けを行って得られた学習用音声データを（1）式のデータ χ^N として用いて記述長を計算することで精度よく記述長を求めることができる。

【0028】

このとき、任意の分布数として、最大分布数をそれぞれの状態に持つHMMを用いることで、より一層、高精度な対応付けが行えるので、そのアライメントデータを記述長の計算に用いることで、より一層、精度よく記述長を求めることができる。

【0029】

また、前記HMMは音節HMMとすることが望ましく、本発明の場合、音節HMMとすることによって演算量の削減などの効果が得られる。たとえば、音節の数を124音節とした場合、音素の数（26から40個程度）に比べると、数の面では音節の方が多いが、音素HMMの場合、トライフォンモデルを音響モデル単位として用いることが多く、このトライフォンモデルは、ある音素の前後の音素環境を考慮して1つの音素として構成されるので、あらゆる組み合わせを考慮すると、そのモデル数は数千個となり、音響モデル数としては音節モデルの方がはるかに少なくなる。

【0030】

ちなみに、音節HMMの場合、それぞれの音節HMMを構成する状態数は子音を含む音節の場合が5個程度、母音だけで構成される音節の場合が3個程度であるので、合計の状態数は約600程度であるが、トライフォンモデルの場合は、状態数の合計は、モデル間で状態共有を行い、状態数を削減した場合であっても数千個にもものぼる。このことから、HMMを音節HMMとすることによって、記述長を求める計算は勿論のこと、全般的な演算量の削減を図ることができ、また、トライフォンモデルに遜色ない認識精度が得られるといった効果が得られる。

【0031】

また、前記HMMが音節HMMである場合、同一子音や同一母音を持つ複数の音節HMMに対し、これらの音節HMMを構成する状態のうち、同一子音を有する音節HMM同士においてはそれら音節HMMにおける初期状態またはこの初期状態を含む少なくとも2つの状態を共有し、同一母音を有する音節HMM同士においてはそれら音節HMMにおける自己ループを有する状態の最終状態またはこの最終状態を含む少なくとも2つの状態を共有するようにしているので、パラメータ数のより一層の削減が可能となり、それによって、演算量の削減、使用メモリ量の削減、処理速度の高速化がより一層図れ、さらに、低価格、低消費電力化の効果もより大きなものとなる。

【0032】

また、本発明の音声認識装置は、上述の本発明の音響モデル作成方法によって作成された音響モデル（HMM）を用いる。すなわち、このHMMはそれを構成する複数の状態ごとに最適な分布数を有した各音節ごとの音節モデルとなっているので、すべての状態が多数の分布数で一定となっているHMMに比べ、認識性能を劣化させることなく、それぞれの音節HMMにおけるパラメータ数を大きく削減することができる。これによって、演算量の削減、使用メモリ量の削減が可能となり、それによって、処理速度の高速化、低価格化、低消費電力化も可能となるので、ハードウェア資源に大きな制約のある小型・安価なシステムに搭載する音声認識装置として極めて有用なものとなる。

【0033】

【発明の実施の形態】

以下、本発明の実施の形態について説明する。

【0034】**〔第1の実施の形態〕**

まず、第1の実施の形態として、それぞれの音節HMMにおいて、MDL基準を用いてその音節HMMを構成するそれぞれの状態ごとに分布数の最適化を行う例について説明する。

【0035】

なお、本発明は音素HMMと音節HMMの両方に適用可能であるが、この第1の実施の形態では音節HMMについて説明する。まず、この第1の実施の形態の全体的な処理の流れの概略について図1により説明する。

【0036】

まず、個々の音節HMMを構成するそれぞれの状態のガウス分布の分布数がある値から最大分布数までに設定した音節HMMセットを作成する。この実施の形態では、分布数は分布数1、分布数2、分布数4、分布数8、分布数16、分布数32、分布数64の7種類の分布数であるとする。

【0037】

すなわち、分布数を1としたすべての音節HMMからなる音節HMMセット、分布数を2としたすべての音節HMMからなる音節HMMセット、分布数を4としたすべての音節HMMからなる音節HMMセットというように、この場合、それぞれの音節について上述の7種類の分布数を有する7種類の音節HMMセットを作成する。なお、この実施の形態では、分布数を7種類として説明するが、7種類に限られるものではなく、また、それぞれの分布数も1, 2, 4, 8, 16, 32, 64というような値に限られるものではなく、また、最大分布数も64に限られるものではない。

【0038】

そして、この7種類の音節HMMセットに含まれるすべての音節HMMに対して、HMM学習部2がそれぞれの音節HMMのパラメータについて最尤推定法を用いてそれぞれ学習し、分布数1から最大分布数までの学習済みの音節HMMが

作成される。すなわち、この実施の形態では、分布数として、分布数1、分布数2、分布数4、・・・、分布数64の7種類としているので、それらに対応した7種類の学習済みの音節HMMセット31～37が作成される。これについて図2により説明する。

【0039】

HMM学習部2では、学習用音声データ1を用いて最尤推定法によってそれぞれの音節（ここでは、音節/a/、音節/ka/、・・・など124音節とする）について分布数を1, 2, ..., 64の7種類とした個々の音節HMMセットの学習を行い、それぞれの分布数ごとの音節HMMセット31, 32, ..., 37を作成する。なお、この例では、それぞれの音節HMMは、自己ループを有する状態がS0, S1, S2の3つの状態で構成されるものとする。

【0040】

これによって、分布数1の音節HMMセット31には、音節/a/のHMM、音節/ka/のHMMなど、124音節それぞれの音節について学習済みの音節HMMが存在し、また、分布数2の音節HMMセット32には、音節/a/のHMM、音節/ka/のHMMなど、124音節それぞれの音節について学習済みの音節HMMが存在するというように、分布数1、分布数2、分布数4、・・・、分布数64のそれぞれの音節HMMセット31, 32, ..., 37には、124音節それぞれの音節について学習済みの音節HMMが存在する。

【0041】

なお、図2において、分布数1の音節HMMセット31、分布数2の音節HMMセット32、・・・、分布数64の音節HMMセット37の各音節HMMの各状態S0, S1, S2の下に描かれている楕円形枠A内のガウス分布がそれぞれの状態における分布例を示すもので、分布数1の音節HMMセット31は、どの音節HMMについても1個の分布を有し、分布数2の音節HMMセット32は、どの音節HMMについても2個の分布を有し、分布数64の音節HMMセット37は、どの音節HMMについても64個の分布を有している。

【0042】

このように、HMM学習部2の学習によって、分布数1の音節HMMセット3

1、分布数2の音節HMMセット32、・・・、最大分布数の音節HMMセット（この場合、分布数64の音節HMMセット37）の7種類の分布数に対応するそれぞれの音節HMMセット31～37が作成される。

【0043】

次に、図1に説明が戻って、HMM学習部2の学習によって学習された分布数1の音節HMMセット31、分布数2の音節HMMセット32、・・・、最大分布数の音節HMMセット（この場合、分布数64の音節HMMセット37）のうち、任意の音節HMMセット（ここでは、最大分布数、つまり、分布数64の音節HMMセット37）を用い、アライメントデータ作成部4によって、すべての学習用音声データ1とのビタビ（Viterbi）アライメントをとり、それぞれの音節HMMの各状態と学習用音声データ1との対応付けを行って、最大分布数（分布数64）の音節HMMセット37の各状態S0、S1、S2と学習用音声データ1とのアライメントデータ5を作成する。これについて図3および図4を参照しながら説明する。

【0044】

なお、図3はこのアライメントデータ作成処理を説明するに必要な部分だけを図1から取り出して示すものであり、また、図4はアライメントデータ作成を作成するために、それぞれの音節HMMの各状態と学習用音声データ1との対応付けを行う処理の具体例を説明するものである。

【0045】

アライメントデータ作成部4では、すべての学習用音声データ1と最大分布数の音節HMMセット（この場合、分布数64の音節HMMセット37）を用いて、図4の（a）、（b）、（c）に示すように、分布数64の音節HMMセット37の各音節HMMにおける各状態S0、S1、S2とその音節に対応する学習用音声データ1とのアライメントをとる。

【0046】

たとえば、図4（b）に示すように、「秋（あき）の・・・」という学習用音声データ例に対してアライメントをとると、その学習用音声データ「あ」、「き」、「の」、・・・に対応する各音声データ区間において、分布数64の音節/a

/のHMMにおける状態S 0は、「あ」の音声データにおける区間 t 1に対応し、音節/a/のHMMにおける状態S 1は「あ」の音声データにおける区間 t 2に対応し、音節/a/のHMMにおける状態S 2は「あ」の音声データにおける区間 t 3に対応するというような対応付けを行って、その対応付けデータをアライメントデータ 5 とする。

【 0 0 4 7 】

同様に、分布数 6 4 の音節/ki/のHMMにおける状態S 0は、「き」の音声データにおける区間 t 4に対応し、音節/ki/のHMMにおける状態S 1は、「き」の音声データにおける区間 t 5に対応し、音節/ki/のHMMにおける状態S 2は、「き」の音声データにおける区間 t 6に対応するというような対応付けを行って、その対応付けデータをアライメントデータ 5 とする。

【 0 0 4 8 】

また、図4 (c) に示すように、学習用音声データの一例として、「試合 (しあい) . . . 」という学習用音声データにおける「し」に対応する部分、「あ」に対応する部分、「い」に対応する部分において、「あ」の部分に注目すると、分布数 6 4 の音節/a/のHMMにおける状態S 0は「あ」の音声データにおける区間 t 1 1に対応し、音節/a/のHMMにおける状態S 1は「あ」の音声データにおける区間 t 1 2に対応し、音節/a/のHMMにおける状態S 2は「あ」の音声データにおける区間 t 1 3に対応するというような対応付けを行って、その対応付けデータをアライメントデータ 5 とする。

【 0 0 4 9 】

次に、このアライメントデータ作成部 4 によって求められた分布数 6 4 の音節HMMセットにおけるそれぞれの音節HMMの各状態と学習用音声データとのアライメントデータ 5 を用いて、分布数 1 から最大分布数までの音節HMMセット (この場合、分布数 1、分布数 2、分布数 4、. . . 、分布数 6 4 の 7 種類の分布数に対応する各音節HMMセット 3 1 ~ 3 7 について、すべての状態の記述長を、図 1 に示す記述長計算部 6 によって求める。これについて図 5 および図 6 を参照しながら説明する。

【 0 0 5 0 】

図5は記述長計算部6の説明に必要な部分を図1から取り出して示すもので、分布数1から最大分布数の各音節HMMセット（この場合、分布数1、分布数2、分布数4、・・・、分布数64の各音節HMMセット31～37のパラメータと、学習用音声データ1と、各音節HMMの各状態と学習用音声データ1とのアライメントデータ5とが記述長計算部6に与えられる。

【0051】

そして、この記述長計算部6によって、各音節HMMにおける各状態のそれぞれの分布数対応の記述長が計算される。これによって、分布数1から最大分布数（分布数64）までの7種類の分布数に対応する各音節HMMセット31～37の各音節HMMにおける各状態の記述長が計算される。

【0052】

すなわち、分布数1の音節HMMセット31の各音節HMMにおける各状態の記述長、分布数2の音節HMMセット32の各音節HMMにおける各状態の記述長、分布数4の音節HMMセット33の各音節HMMにおける各状態の記述長、分布数64の音節HMMセット37の各音節HMMにおける各状態の記述長というように、分布数1の音節HMMセット31の各音節HMMにおける各状態の記述長から分布数64の各音節HMMにおける各状態の記述長が得られ、これら、分布数1の音節HMMセット31の各音節HMMにおける各状態の記述長71から分布数64の各音節HMMにおける各状態の記述長は、記述長格納部71～77に保持される。なお、この記述長の計算の仕方については後に説明する。

【0053】

図6は図5で求められた分布数1の音節HMMの各音節HMMにおける各状態の記述長（記述長格納部71に保持されている各状態の記述長）から最大分布数（分布数64）の音節HMMセットの各音節HMMにおける各状態の記述長（記述長格納部77に保持されている各状態の記述長）において、たとえば、音節/a/のHMMの各状態S0、S1、S2についてそれぞれ記述長が求められた様子を示すものである。

【0054】

この図6からもわかるように、分布数1における音節/a/のHMMの状態S0

、S1、S2についてそれぞれ記述長が求められ、分布数2における音節/a/のHMMの状態S0、S1、S2についてそれぞれ記述長が求められ、分布数64における音節/a/のHMMの状態S0、S1、S2についてそれぞれ記述長が求められるというように、分布数1から最大分布数（分布数64）までの7種類の分布数に対応する音節/a/のHMMについて、それぞれの状態S0、S1、S2の記述長が求められる。なお、この図6では、7種類の分布数のうち分布数1と最大分布数（分布数64）の音節/a/のHMMについてのみが図示されている。

【0055】

そのほかの音節についても同様に、分布数1から最大分布数（分布数64）までの7種類の分布数に対応するそれぞれの音節HMMについて、それぞれの状態S0、S1、S2ごとに記述長が求められる。

【0056】

次に、状態選択部8が上述の記述長計算部6で計算された分布数1の音節HMMセット31の各状態の記述長から最大分布数（分布数64）の音節HMMセット37の各状態の記述長を用い、各音節HMMごとに、各音節HMMの各状態の記述長が最小となる分布数を持つ状態を選択する。これを図7および図8を参照しながら説明する。

【0057】

図7は状態選択部8の説明に必要な部分を図1から取り出して示すもので、記述長計算部6で計算された分布数1の音節HMMセット31の各状態の記述長（記述長格納部71に保持されている各状態の記述長）から最大分布数（分布数64）の音節HMMセット37の各状態の記述長（記述長格納部77に保持されている各状態の記述長）について、それぞれの音節HMMごとにそれぞれの状態S0、S1、S2において、どの分布数を持つ状態の記述長が最小となるかを判断し、記述長が最小となる分布数を持つ状態を選択する。

【0058】

ここでは、音節/a/のHMMと音節/ka/のHMMについて、分布数1から最大分布数（分布数64）までの7種類の分布数に対応するそれぞれの音節HMMにおけるそれぞれの状態S0、S1、S2ごとに、どの分布数を持つ状態の記述長

が最小（記述長最小）となるかを判断し、記述長が最小となる分布数を持つ状態の選択処理を図8によって説明する。

【0059】

まず、音節/a/のHMMにおける状態S0について、分布数1から分布数64の中でどの分布数を持つ状態S0が記述長最小であるかを判断した結果、分布数2を持つ状態S0が記述長最小であると判断されたとする。これを点線の矩形枠M1で示す。

【0060】

また、音節/a/のHMMにおける状態S1について、分布数1から分布数64の中でどの分布数を持つ状態S1が記述長最小であるかを判断した結果、分布数64を持つ状態S1が記述長最小であると判断されたとする。これを点線の矩形枠M2で示す。

【0061】

また、音節/a/のHMMにおける状態S2について、分布数1から分布数64の中でどの分布数を持つ状態S2が記述長最小であるかを判断した結果、分布数1を持つ状態S2が記述長最小であると判断されたとする。これを点線の矩形枠M3で示す。

【0062】

このように、この音節/a/のHMMについて、分布数1から最大分布数（分布数64）までのそれぞれの状態S0、S1、S2ごとに、どの分布数を持つ状態の記述長が最小となるかを判断し、記述長最小を持つ状態を選択すると、この場合、状態S0にあっては分布数2を持つ状態S0が選択され、状態S1にあっては分布数64を持つ状態S1が選択され、状態S2にあって分布数1を持つ状態S2が選択されるので、それらを結合した音節/a/のHMMを構築する。

【0063】

この記述長最小を持つ状態で構成された音節/a/のHMMは、その状態S0は分布数が2、状態S1は分布数が64、状態S2は分布数が1となり、分布数が最適化された状態の結合による音節/a/のHMMとなる。

【0064】

同様に、音節/ka/のHMMにおける状態S0について、分布数1から分布数64の中でどの分布数を持つ状態S0が記述長最小かを判断した結果、分布数1を持つ状態S0が記述長最小であると判断されたとする。これを点線の矩形枠M4で示す。

【0065】

また、音節/ka/のHMMにおける状態S1について、分布数1から分布数64の中でどの分布数を持つ状態が記述長最小かを判断した結果、分布数2を持つ状態S1が記述長最小であると判断されたとする。これを点線の矩形枠M5で示す。また、音節/ka/のHMMにおける状態S2について、分布数1から分布数64の中でどの分布数を持つ状態S2が記述長最小かを判断した結果、同じく、分布数2を持つ状態S2が記述長最小であると判断されたとする。これを点線の矩形枠M6で示す。

【0066】

このように、この音節/ka/のHMMについて、分布数1から最大分布数（分布数64）までのそれぞれの状態S0、S1、S2ごとに、どの分布数を持つ状態の記述長が最小となるかを判断し、記述長最小を持つ状態を選択すると、この場合、状態S0にあっては分布数1を持つ状態S0が選択され、状態S1にあっては分布数2を持つ状態が選択され、状態S2は分布数2を持つ状態S3が選択されるので、それらを結合した音節/ka/のHMMを構築する。

【0067】

この記述長最小を持つ状態で構成された音節/ka/のHMMは、状態S0は分布数が1、状態S1は分布数が2、状態S2も分布数が2となり、分布数が最適化された状態の結合による音節/ka/のHMMとなる。

【0068】

このような処理をすべての音節（ここでは124音節）のHMMについて行うことによって、それぞれの音節HMMは、記述長最小を持つ状態で構成され、それによって、最適化された分布数を持つHMMが構築される。

【0069】

このようにして、それぞれの音節HMMについて、各状態ごとに最適化された

分布数を持つHMMが構築されると、HMM再学習部9（図1参照）によって、これら最適化された分布数を持つHMMの全パラメータに対し、学習用音声データ1を用いて最尤推定法によって再学習する。これによって、それぞれの音節HMMについて、各状態ごとに最適化された分布数を持ち、かつ、それぞれの状態ごとに最適なパラメータが得られた音節HMMセット10が得られる。

【0070】

次に、本発明で用いるMDL（記述長最小）基準について説明する。このMDL基準については、たとえば、「韓太舜著“岩波講座応用数学11、情報と符号化の数理” 岩波書店（1994）、pp249-275」などに述べられている公知の技術であり、従来技術の項でも述べたように、モデルの集合 $\{1, \dots, i, \dots, I\}$ とデータ $\chi^N = \{\chi_1, \dots, \chi_N\}$ （ただし、Nはデータ長）が与えられたときのモデル i を用いた記述長 $l_i(\chi^N)$ は、前述した（1）式のように定義され、このMDL基準は、この記述長 $l_i(\chi^N)$ が最小であるモデルが最適なモデルであるとしている。

【0071】

本発明では、ここでいうモデル集合 $\{1, \dots, i, \dots, I\}$ は、あるHMMにおいて分布数がある値から最大分布数までの複数種類に設定されたある状態の集合であるとして考える。なお、分布数がある値から最大分布数までの複数種類に設定されているときの分布数の種類が I 種類（ I は $I \geq 2$ の整数）であるとしたとき、上述の $1, \dots, i, \dots, I$ は、1番目の種類から I 番目の種類までそれぞれの種類を特定するための符号であって、上述の（1）式を、 $1, \dots, i, \dots, I$ のうちの i 番目の分布数の種類を持つ状態の記述長を求める式として用いるものである。

【0072】

なお、この $1, \dots, i, \dots, I$ の I は、異なる分布数を持つHMMセットの総数、すなわち、分布数が何種類あるかを表すもので、この実施の形態では、分布数は、1, 2, 4, 8, 16, 32, 64の7種類としているので、 $I = 7$ となる。

【0073】

このように、 $1, \dots, i, \dots, I$ が、1 番目の種類から I 番目の種類までそれぞれの種類を特定するための符号であるので、この実施の形態での例では、分布数 1 に対しては分布数の種類を表す符号として、 $1, \dots, i, \dots, I$ のうち 1 が与えられ、分布数の種類が 1 番目であることを示す。また、分布数 2 に対しては分布数の種類を表す符号として、 $1, \dots, i, \dots, I$ のうち 2 が与えられ、分布数の種類が 2 番目であることを示す。また、分布数 4 に対しては分布数の種類を表す符号として、 $1, \dots, i, \dots, I$ のうち 3 が与えられ、分布数の種類が 3 番目であることを示す。また、分布数 8 に対しては分布数の種類を表す符号として、 $1, \dots, i, \dots, I$ のうち 4 が与えられ、分布数の種類が 4 番目であることを示す。また、分布数 16 に対しては分布数の種類を表す符号として、 $1, \dots, i, \dots, I$ のうち 5 が与えられ、分布数の種類が 5 番目であることを示す。また、分布数 32 に対しては分布数の種類を表す符号として、 $1, \dots, i, \dots, I$ のうち 6 が与えられ、分布数の種類が 6 番目であることを示す。また、分布数 64 に対しては分布数の種類を表す符号として、 $1, \dots, i, \dots, I$ のうち 7 が与えられ、分布数の種類が 7 番目であることを示す。

【0074】

ここで、音節/a/のHMMについて考えると、図8に示すように、分布数1から分布数64までの7種類の分布数をもつ状態S0の集合が1つのモデル集合、同じく、分布数1から分布数64までの7種類の分布数をもつ状態S1の集合が1つのモデル集合、同じく、分布数1から分布数64までの7種類の分布数をもつ状態S2の集合が1つのモデル集合となる。

【0075】

したがって、上述の(1)式のように定義された記述長 $l_i(\chi^N)$ は、本発明においては、ある状態の分布数の種類が $1, \dots, i, \dots, I$ のうちの i 番目の種類に設定したときのその状態(これを状態 i で表す)の記述長 $l_i(\chi^N)$ であるとして、次式のように定義する。

【0076】

【数 2】

$$l_i(x^N) = -\log P_{\hat{\theta}(i)}(x^N) + \alpha \left(\frac{\beta_i}{2} \log N \right) \quad (2)$$

$\hat{\theta}(i)$: 状態 i のパラメータ $\theta^{(i)} = \theta_1^{(i)}, \dots, \theta_{\beta_i}^{(i)}$ の最尤推定量

【0077】

この(2)式は、前述の(1)式における右辺の最終項である第3項の $\log I$ は定数であるので省略し、かつ、(1)式における右辺の第2項である $(\beta_i / 2) \log N$ に重み係数 α を乗じている点が(1)式と異なっている。なお、上述の(2)式においては、(1)式における右辺の最終項である第3項の $\log I$ を省略したが、これを省略せずにそのまま残した式としてもよい。

【0078】

また、 β_i は分布数の種類が i 番目の分布数を持つ状態 i の次元 (自由度) として、分布数 \times 特徴ベクトルの次元数で表されるが、この特徴ベクトルの次元数は、ここでは、ケプストラム (CEP) 次元数 $+$ Δ ケプストラム (CEP) 次元数 $+$ Δ パワー (POW) 次元数である。

【0079】

また、 α は最適な分布数を調整するための重み係数であり、この α を変えることによって、記述長 $l_i(x^N)$ を変化させることができる。すなわち、図9(a), (b) に示すように、単純に考えれば、(2)式の右辺の第1項は、分布数の増加に伴ってその値が減少し (細い実線で示す)、(2)式における右辺の第2項は、分布数の増加に伴って単調増加 (太い実線で示す) し、これら第1項と第2項の和で求められる記述長 $l_i(x^N)$ は、破線で示すような値をとる。

【0080】

したがって、 α を可変することによって、第2項の単調増加の傾きを可変 (α を大きくするほど傾きが大きくなる) することができるので、(2)式における右辺の第1項と第2項の和で求められる記述長 $l_i(x^N)$ は、 α の値を変化さ

せることによって変化させることができる。これによって、たとえば、 α を大きくすると、図9 (a) は同図 (b) のようになり、分布数がより小さい場合に記述長 $l_i (\chi^N)$ が最小になるように調整することができる。

【0081】

なお、(2) 式における分布数の種類が i 番目の分布数を持つ状態 i は M 個のデータ (あるフレーム数からなる M 個のデータ) に対応している。すなわち、データ 1 の長さ (フレーム数) を n_1 、データ 2 の長さ (フレーム数) を n_2 、データ M の長さ (フレーム数) を n_M で表せば、 χ^N の N は $N = n_1 + n_2 + \dots + n_M$ で表されるので、(2) 式における右辺の第 1 項は、下記の (3) 式のように表される。

【0082】

なお、ここでのデータ 1, データ 2, \dots , データ M は、状態 i に対応つけられた多数の学習用音声データ 1 のある区間に対応するデータ (たとえば、図 4 で説明したように、仮に状態 i が分布数 64 の音節/a/の HMM における状態 S_0 であるとすれば、区間 t_1 や区間 t_{11} に対応する学習用音声データ) である。

【0083】

【数 3】

$$\log P_{\theta(i)} (x^N) = \log P_{\theta(i)} (x^{n_1}) + \log P_{\theta(i)} (x^{n_2}) + \dots + \log P_{\theta(i)} (x^{n_M}) \quad (3)$$

【0084】

この (3) 式において、右辺のそれぞれの項は、分布数の種類が i 番目の分布数を持つ状態 i に対応する区間のデータに対する尤度であるが、この実施の形態では、当該状態 i に対応する区間のデータに対する出力確率としている。なお、その出力確率は、実際には、その状態 i に対応するデータを構成する複数のフレーム対応の出力確率の和で表される。

【0085】

ところで、上述の(2)式によって求められる記述長 $l_i(\chi^N)$ において、記述長 $l_i(\chi^N)$ が最小であるモデルが最適なモデル、すなわち、ある音節 HMM のある状態において、記述長 $l_i(\chi^N)$ が最小となる分布数を持つ状態が最適な状態であるとする。

【0086】

すなわち、この実施の形態では、分布数を 1, 2, 4, 8, 16, 32, 64 の 7 種類としているので、記述長 $l_i(\chi^N)$ は、ある状態において、分布数 1 (分布数の種類としては 1 番目) としたときの当該状態の記述長 $l_1(\chi^N)$ 、分布数 2 (分布数の種類としては 2 番目) としたときの当該状態の記述長 $l_2(\chi^N)$ 、分布数 4 (分布数の種類としては 3 番目) としたときの当該状態の記述長 $l_3(\chi^N)$ 、分布数 8 (分布数の種類としては 4 番目) としたときの記述長 $l_4(\chi^N)$ 、分布数 16 (分布数の種類としては 5 番目) のときの記述長 $l_5(\chi^N)$ 、分布数 32 (分布数の種類としては 6 番目) のときの当該状態の記述長 $l_6(\chi^N)$ 、分布数 64 (分布数の種類としては 7 番目) としたときの当該状態の記述長 $l_7(\chi^N)$ の 7 種類の記述長が得られ、その中から記述長が最小となる分布数を持つ状態 i を選択する。

【0087】

たとえば、図 8 の例においては、音節/a/の HMM について考えると、分布数 1 から最大分布数 (分布数 64) までのそれぞれの状態 S_0 , S_1 , S_2 ごとに、それぞれの分布数を持つ状態の記述長を (2) 式によって計算して求め、記述長最小の状態を選択すると、この図 8 は、前述したように、状態 S_0 にあっては分布数 2 の状態 S_0 が記述長最小であるとしてこの分布数 2 の状態 S_0 が選択され、状態 S_1 にあっては分布数 64 の状態 S_1 が記述長最小であるとしてこの分布数 64 の状態 S_1 が選択され、状態 S_2 にあっては分布数 1 の状態 S_2 が記述長最小であるとしてこの分布数 1 の状態 S_2 が選択された例である。

【0088】

以上説明したように、(2)式を用いて、それぞれの音節 HMM について、分布数 1 から最大分布数 (この実施の形態では分布数 64) までのそれぞれの状態 (この実施の形態では状態 S_0 , S_1 , S_2) ごとに、記述長 $l_i(\chi^N)$ を計

算し、それぞれの状態において、どの分布数を持つ状態の記述長が最小となるかを判断し、記述長最小となった状態を選択する。そして、それぞれの音節ごとに、記述長最小となる分布数を持つ状態によってその音節HMMを構築する。

【0089】

このようにして、それぞれの音節HMMについて、各状態ごとに最適化された分布数を持つHMMが構築されると、これらのHMMの全パラメータに対し、学習用音声データ1を用いて最尤推定法によって再学習する。これによって、それぞれの音節HMMについて、各状態ごとに最適化された分布数を持ち、かつ、それぞれの状態ごとに最適なパラメータが得られる。

【0090】

この各状態ごとに最適化された分布数を持ち、かつ、それぞれの状態ごとに最適なパラメータが得られた各音節HMMは、各音節HMMにおいて各状態ごとに分布数が最適化されているため、十分な認識性能を確保することができ、しかも、すべての状態で同じ分布数とした場合に比べ、パラメータ数を大幅に削減することができ、演算量の削減、使用メモリ量の削減が図れ、処理速度の高速化が図れ、さらに、低価格、低消費電力化も可能となる。

【0091】

図10はこのようにして作成された音響モデル（HMMモデル）を用いた音声認識装置の構成を示す図であり、音声入力用のマイクロホン21、このマイクロホン21から入力された音声を増幅するとともにディジタル信号に変換する入力信号処理部22、入力信号処理部からのディジタル変換された音声信号から特徴データ（特徴ベクトル）を抽出する特徴分析部23、この特徴分析部23から出力される特徴データに対し、HMMモデル24や言語モデル25を用いて音声認識する音声認識処理部26から構成され、このHMMモデル24として、これまで説明した音響モデル作成方法によって作成されたHMMモデル（図1で示した状態ごとに最適な分布数を持つ音節HMMセット10）を用いる。

【0092】

このように、この音声認識装置はそれぞれの音節HMM（たとえば、124音節ごとの音節HMM）において、その音節HMMを構成するそれぞれの状態ごと

に最適な分布数を有した音節モデルとなっているので、高い認識性能を維持した上で、それぞれの音節HMMにおけるパラメータ数を大きく削減することができ、それによって、演算量の削減、使用メモリ量の削減が図れ、処理速度の高速化が図れ、さらに、低価格、低消費電力化も可能となるので、ハードウェア資源に大きな制約のある小型・安価なシステムにも搭載する音声認識装置として極めて有用なものとなる。

【0093】

ちなみに、本発明の状態ごとに最適な分布数を持つ音節HMMセット10を用いた音声認識装置を用いた認識実験として、124音節HMMにおける文の認識実験を行ったところ、総分布数が約19000での認識率が94.6%であったものを、本発明によって分布数の最適化を行い、総分布数を約7000としたときの認識率が94.4%となり、総分布数数を約1/3としても認識性能を維持できることが確認できた。

【0094】

〔第2の実施の形態〕

この第2の実施の形態では、同一子音や同一母音を持つ音節HMMにおいて、これらの音節HMMを構成する複数の状態（自己ループを有する状態）のうち、たとえば、初期状態または最終状態を共有した音節HMM（これをここでは便宜的に状態共有音節HMMと呼ぶことにする）を構築し、その状態共有音節HMMに対して、前述の第1の実施の形態で説明した技術、すなわち、それぞれの音節HMMの各状態の分布数を最適化する技術を適用する。以下、図11を参照しながら説明する。

【0095】

ここでは、同一子音や同一母音を持つ音節HMMとして、たとえば、音節/ki/のHMM、音節/ka/のHMM、音節/sa/のHMM、音節/a/のHMMについて考える。すなわち、音節/ki/と音節/ka/はともに子音/k/を持ち、音節/ka/、音節/sa/、音節/a/はともに母音/a/を持っている。

【0096】

そこで、同一子音を持つ音節HMMにおいては、それぞれの音節HMMにおい

て、前段に存在する状態（ここでは、第1の状態とする）を共有し、同一母音を持つ音節HMMにおいては、それぞれの音節HMMにおいて、後段に存在する状態（ここでは、自己ループを有する状態のうち最終状態とする）を共有する。

【0097】

図11は、音節/ki/のHMMの第1状態S0と音節/ka/のHMMの第1状態S0とを共有し、音節/ka/のHMMの最終状態S4と音節/sa/のHMMの自己ループを有する最終状態S4と音節/a/のHMMの自己ループを有する最終状態S2をそれぞれ共有することを表す図であり、それぞれ共有する状態を太い実線で示す楕円枠Cで囲っている。

【0098】

このように、同一子音や同一母音を持つ音節HMMにおいて、状態共有がなされ、その状態共有された状態は、そのパラメータも同一となり、HMM学習（最尤推定）を行う際に同じパラメータとして扱われる。

【0099】

たとえば、図12に示すように、「かき」という音声データに対し、自己ループを有する状態がS0, S1, S2, S3, S4の5つの状態でなる音節/ka/のHMMと、同じく自己ループを有する状態がS0, S1, S2, S3, S4の5つの状態でなる音節/ki/のHMMとが連結されたHMMが構築されたとき、音節/ka/のHMMの第1の状態S0と音節/ki/のHMMの第1の状態S0が共有されることによって、これら音節/ka/のHMMの状態S0と音節/ki/のHMMの状態S0はそれぞれのパラメータが同一として扱われて同時に学習される。

【0100】

このような状態共有がなされることによって、パラメータ数が減少し、それによって、使用メモリ量の削減、演算量の削減が図れ、処理能力の低いCPUでの動作が可能となり、低消費電力化も図れるので、低価格が要求されるシステムへの適用が可能となる。また、学習用音声データの少ない音節では、パラメータ数の削減によって、過学習による認識性能劣化を防ぐ効果も期待できる。

【0101】

このようにして状態共有がなされることによって、ここでの例で取り上げた音

節/ki/のHMMと音節/ka/のHMMにおいては、それぞれの第1状態S0を共有したHMMが構築される。また、音節/ka/のHMMと音節/sa/のHMMと音節/a/のHMMにおいては、最終状態（図11の例では、音節/ka/のHMMの状態S44と音節/sa/のHMMの状態S4、音節/a/のHMMの状態S2）を共有したHMMが構築される。

【0102】

そして、このように状態共有したそれぞれの音節HMMについて、前述の第1の実施の形態で説明したMDL基準を用いてそれぞれの状態ごとに分布数の最適化を行う。

【0103】

このように、この第2の実施の形態では、同一子音や同一母音を持つ音節HMMにおいて、これらの音節HMMを構成する複数の状態のうち、たとえば、第1状態または最終状態を共有した状態共有音節HMMを構築し、その状態共有音節HMMに対して、前述の第1の実施の形態で説明した技術を適用することによって、パラメータのより一層の削減が図れ、それによって、演算量の削減、使用メモリ量の削減、処理速度の高速化がより一層図れ、さらに、低価格、低消費電力化の効果もより大きなものとなる。さらに、各状態ごとに最適化された分布数を持ち、かつ、それぞれの状態ごとに最適なパラメータが得られた音節HMMとすることができる。

【0104】

したがって、このように状態共有され、かつ、その状態共有されたそれぞれの音節HMMに対して、前述の第1の実施の形態で説明したように、各状態ごとに最適な分布数を持つ音節HMMを作成し、それを図10に示すような音声認識装置に適用することで、高い認識性能を維持した上で、それぞれの音節HMMにおけるパラメータ数をより一層削減することができる。これによって、演算量や使用メモリ量のより一層の削減が図れ、処理速度の高速化が図れ、さらに、低価格、低消費電力化も可能となるので、低コストが要求されハードウェア資源に大きな制約のある小型・安価なシステムにも搭載する音声認識装置として極めて有用なものとなる。

【0105】

なお、上述の状態共有の例では、同一子音や同一母音を持つ音節HMMにおいて、これらの音節HMMを構成する複数の状態のうち、初期状態と最終状態をそれぞれ共有する例について説明したが、それぞれ複数ずつの状態を共有するようにしてもよい。すなわち、同一子音を有する音節HMM同士においては、それら音節HMMにおける初期状態またはこの初期状態を含む少なくとも2つの状態（たとえば、初期状態と第2状態）を共有し、同一母音を有する音節HMM同士においてはそれら音節HMMにおける自己ループを有する状態の最終状態またはこの最終状態を含む少なくとも2つの状態（たとえば、最終状態とそれより1つ手前の状態）を共有するそれによって、パラメータ数をより一層削減することができる。

【0106】

図13は前述した図11において、音節/ki/のHMMの初期状態である第1状態S0および第2状態S1と音節/ka/のHMMの初期状態である第1状態S0および第2の状態S1とをそれぞれ共有し、音節/ka/のHMMの最終状態S4およびそれより1つ前の第4状態S3と音節/sa/のHMMの最終状態S4およびそれよりも1つ前の状態S3と音節/a/のHMMの最終状態S2およびそれよりも1つ前の状態S1をそれぞれ共有することを示した図であり、この図13においてもそれぞれ共有する状態を太い実線で示す楕円枠Cで囲っている。

【0107】

なお、本発明は上述の実施の形態に限られるものではなく、本発明の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、前述の第2の実施の形態では、音節HMMを連結する際、同一子音や同一母音については状態を共有することについて説明したが、たとえば、音素HMMを連結して音節HMMを構築するような場合、同じような考え方で、同一母音についてはその状態の分布を共有することも可能である。

【0108】

たとえば、図14に示すように、音素/k/のHMMと音素/s/のHMMと音素/a/のHMMがあつて、音素/k/のHMMと音素/a/のHMMを連結して音節/ka/の

HMMを構築し、また、音素/s/のHMMと音素/a/のHMMを連結して音節/sa/のHMMを構築する際、新たに構築された音節/ka/のHMMと音節/sa/のHMMの母音/a/は同じであるので、その音節/ka/のHMMと音節/sa/のHMMにおける音素/a/に対応する部分は、音素/a/のHMMの各状態における分布を共有する。

【0109】

そして、このように同一母音の分布を共有した音節/ka/のHMMと音節/sa/のHMMについて第1の実施の形態で説明した状態ごとの分布数の最適化を行うが、この最適化の結果、分布を共有した音節HMM（図14の例では、音節/ka/のHMMと音節/sa/のHMM）においては、その分布共有部分（この図14の例では、音素/a/の自己ループを有する状態）の分布数は音節/ka/のHMMと音節/sa/のHMMで同じとする。

【0110】

このように、分布を共有することで、それぞれの音節HMMにおけるパラメータ数をより一層削減することができ、それによって、演算量や使用メモリ量のより一層の削減が図れるなど、前述の状態共有の場合と同様の効果が得られる。

【0111】

また、本発明は以上説明した本発明を実現するための処理手順が記述された処理プログラムを作成し、その処理プログラムをフロッピーディスク、光ディスク、ハードディスクなどの記録媒体に記録させておくこともでき、本発明は、その処理プログラムの記録された記録媒体をも含むものである。また、ネットワークから当該処理プログラムを得るようにしてもよい。

【0112】

【発明の効果】

以上説明したように本発明の音響モデル作成方法によれば、それぞれの状態ごとにガウス分布数の最適化を行うために、HMMを構成する複数の状態ごとに、分布数がある値から最大分布数まで設定し、この分布数がある値から最大分布数まで設定された状態に対して、分布数がある値から最大分布数のどの分布数が最適であるかを記述長最小基準を用いて選択し、記述長が最小となる分布数を持つ

状態によってそれぞれのHMMを構築し、その構築されたそれぞれのHMMに対して学習用音声データを用いて再学習するようにしている。これによって、少ない演算量で最適な分布数の設定が可能となり、少ない演算量で高い認識性能が得られるHMMを作成することができる。

【0113】

特に、本発明の場合、分布数がある値から最大分布数までの中から最適な分布数を持つ状態を選択するというものであるため、たとえば、ある状態ごとの分布数の種類を7種類とすれば、1つの状態において記述長を求める計算を7回行って、その中から記述長最小となる状態を選択すればよいので、少ない演算量で最適な分布数の設定が可能となる。

【0114】

また、本発明の音声認識装置は、本発明の音声認識装置は、上述の本発明の音響モデル作成方法によって作成された音響モデル（HMM）用いている。すなわち、このHMMはそれを構成する複数の状態ごとに最適な分布数を有した各音節ごとの音節モデルとなっているので、すべての状態が多数の分布数で一定となっているHMMに比べ、認識性能を劣化させることなく、それぞれの音節HMMにおけるパラメータ数を大きく削減することができる。これによって、演算量の削減、使用メモリ量の削減が可能となり、それによって、処理速度の高速化、低価格化、低消費電力化も可能となるので、ハードウェア資源に大きな制約のある小型・安価なシステムに搭載する音声認識装置として極めて有用なものとなる。

【図面の簡単な説明】

【図1】 本発明の第1の実施の形態における音響モデル作成手順を説明する図である。

【図2】 分布数を1から最大分布数（分布数64）までの7種類としたときの音節HMMセット作成について説明する図である。

【図3】 図1で示した音響モデル作成処理においてアライメントデータ作成処理を説明するに必要な部分だけを図1から取り出して示す図である。

【図4】 アライメントデータ作成を作成するために、それぞれの音節HMMの各状態と学習用音声データ1との対応付けを行う処理の具体例を説明する図

である。

【図 5】 図 1 で示した音響モデル作成処理において分布数 1 から最大分布数の各音節 HMM における各状態の記述長を求める処理を説明するに必要な部分だけを図 1 から取り出して示す図である。

【図 6】 音節/a/の HMM において分布数 1 から最大分布数における各状態の記述長が求められた様子を示す図である。

【図 7】 図 1 で示した音響モデル作成処理において MDL 基準による状態選択を説明するに必要な部分だけを図 1 から取り出して示す図である。

【図 8】 MDL 基準によって分布数 1 から最大分布数までのそれぞれの音節 HMM におけるそれぞれの状態 S_0 , S_1 , S_2 ごとに記述長が最小となる状態を選択する処理を説明する図である。

【図 9】 この第 1 の実施の形態で用いる重み係数 α について説明する図である。

【図 10】 本発明の音声認識装置の概略的な構成を説明する図である。

【図 11】 本発明の第 2 の実施の形態である状態共有について説明する図であり、いくつかの音節 HMM において初期状態または最終状態（自己ループを有する状態の中での最終状態）を共有する場合を説明する図である。

【図 12】 初期状態を状態共有した 2 つの音節 HMM を連結したものをある音声データに対応つけて示す図である。

【図 13】 本発明の第 2 の実施の形態である状態共有について説明する図であり、いくつかの音節 HMM において初期状態および第 2 状態または最終状態（自己ループを有する状態の中での最終状態）およびそれより 1 つ前の状態を共有する場合を説明する図である。

【図 14】 本発明のその他の実施の形態として、分布共有について説明する図であり、子音の音素 HMM と母音の音素 HMM を連結して音節 HMM を構築する際、母音の HMM の状態の分布数を共有する場合を説明する図である。

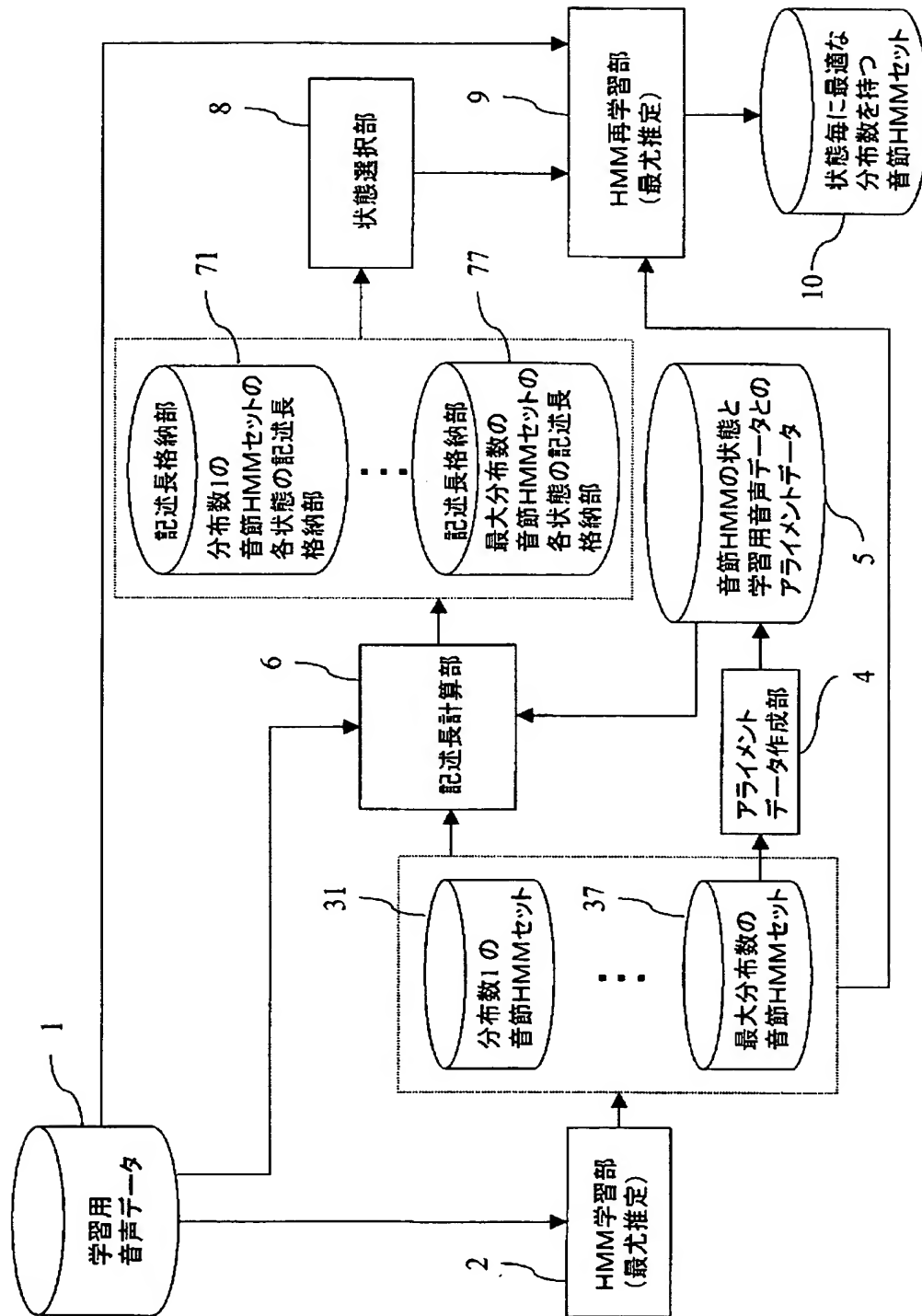
【符号の説明図】

- 1 学習用音声データ
- 2 HMM 学習部

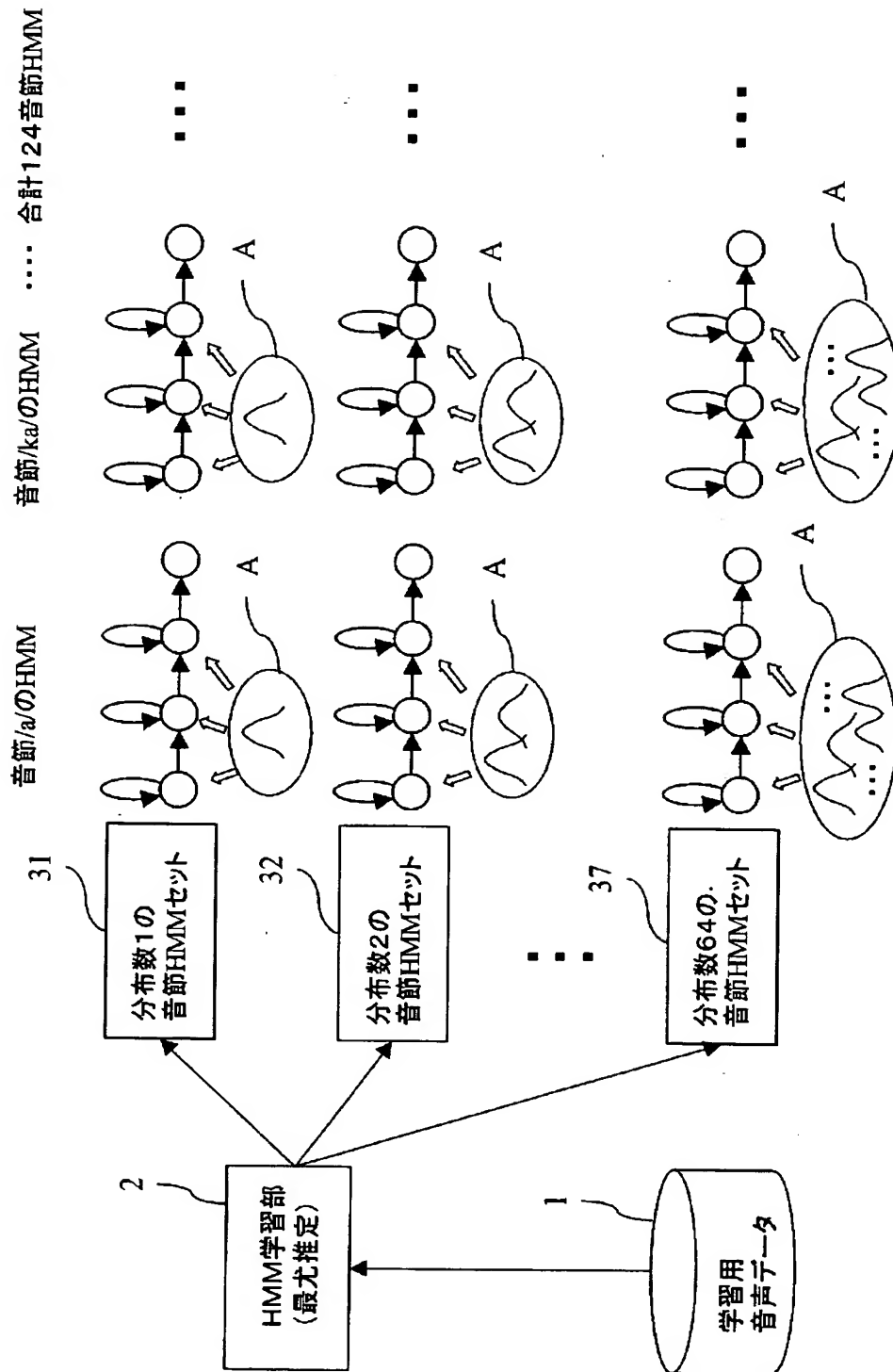
- 31 ~ 37 分布数1から最大分布数の音節HMMセット
- 4 アライメントデータ作成部
- 5 音節HMMの状態と学習用音声データとのアライメントデータ
- 6 記述長計算部
- 71 ~ 77 記述長格納部
- 8 状態選択部
- 9 HMM再学習部
- 10 状態ごとに最適な分布数を持つ音節HMMセット
- 21 マイクロホン
- 22 入力信号処理部
- 23 特徴分析部
- 24 HMMモデル
- 25 言語モデル
- 26 音声認識処理部
- S0, S1, S2, . . . 状態

【書類名】 図面

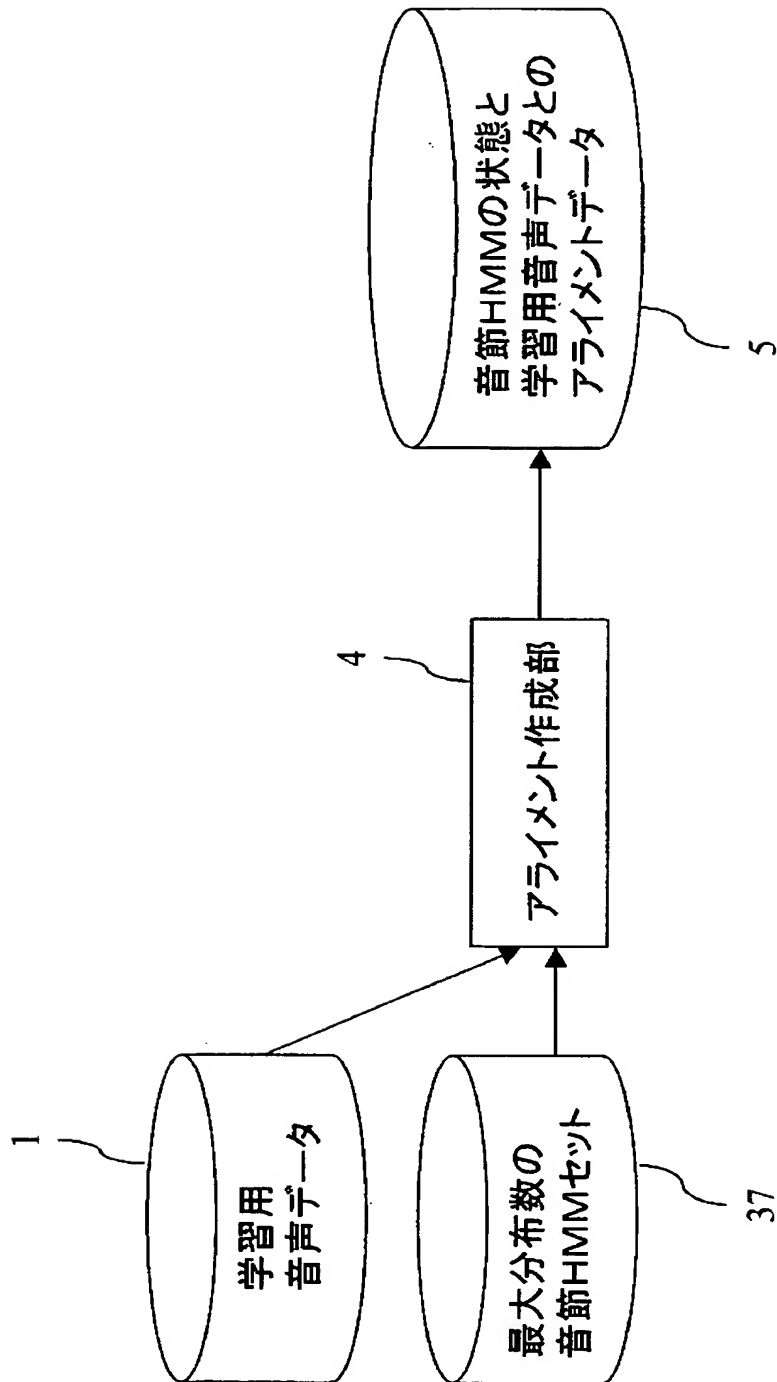
【図1】



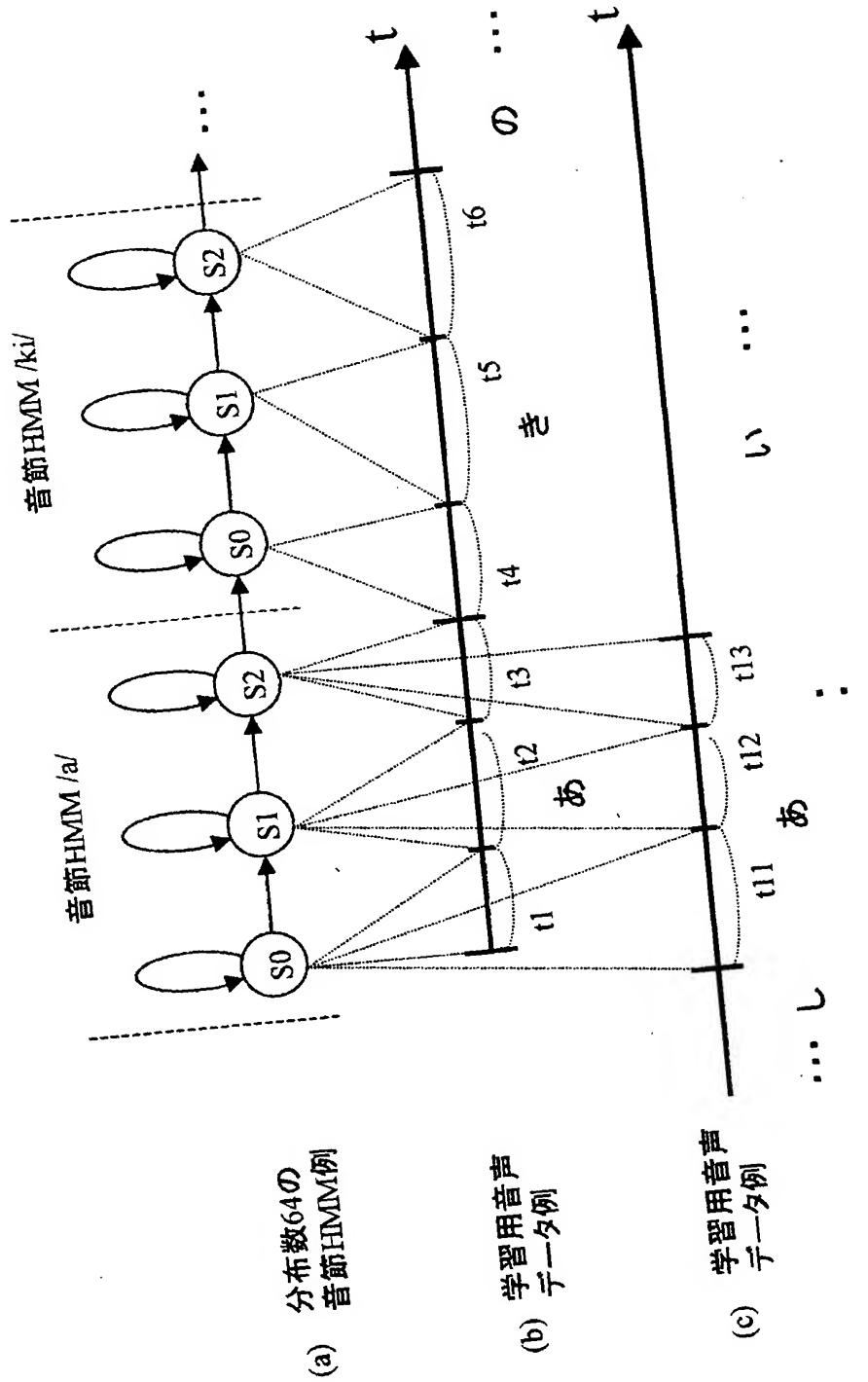
【図2】



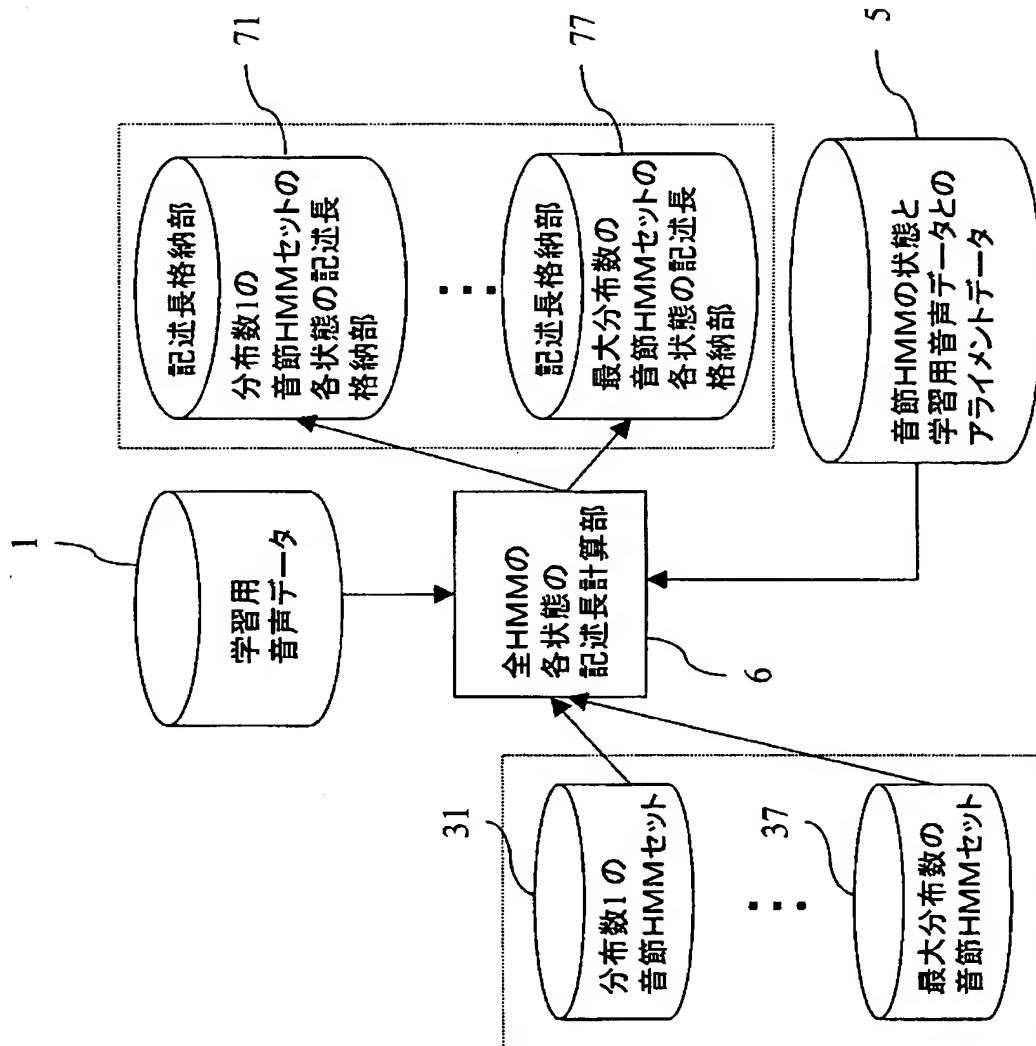
【図 3】



【図4】

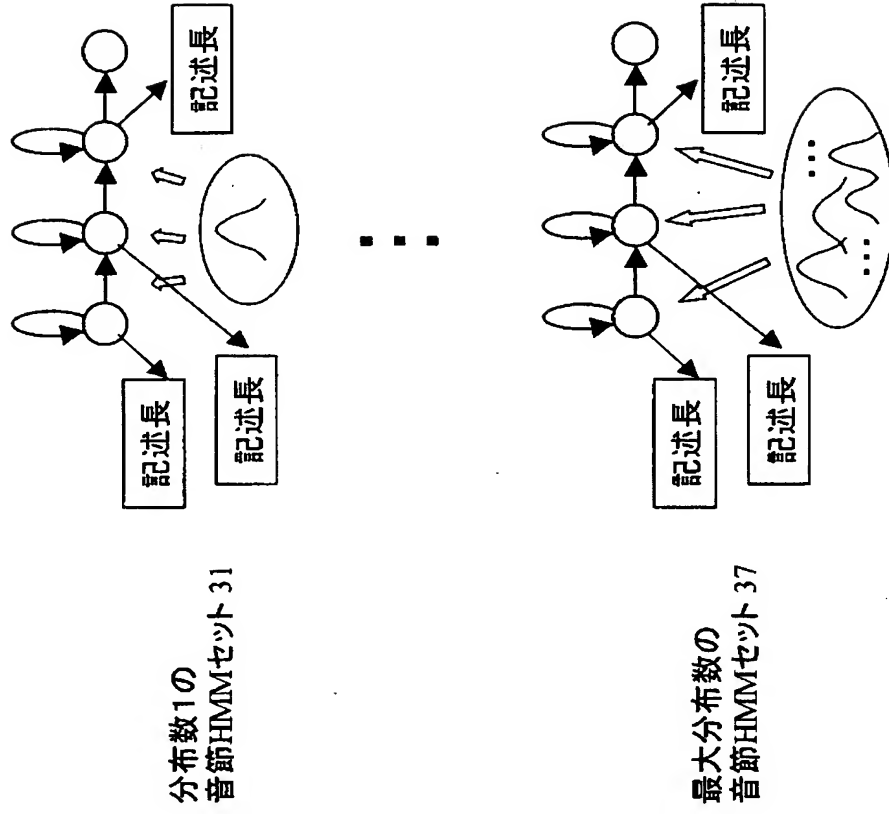


【図 5】

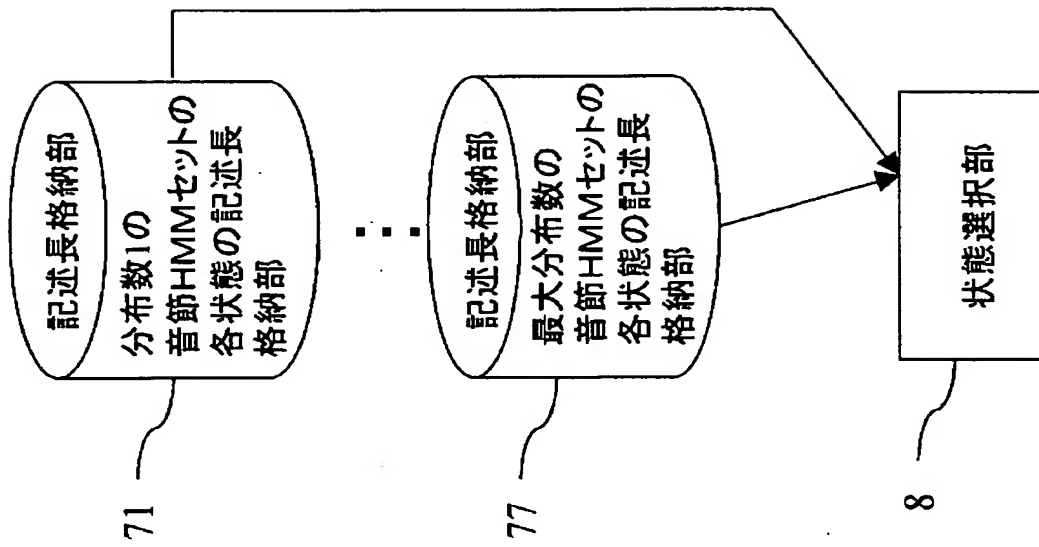


【図 6】

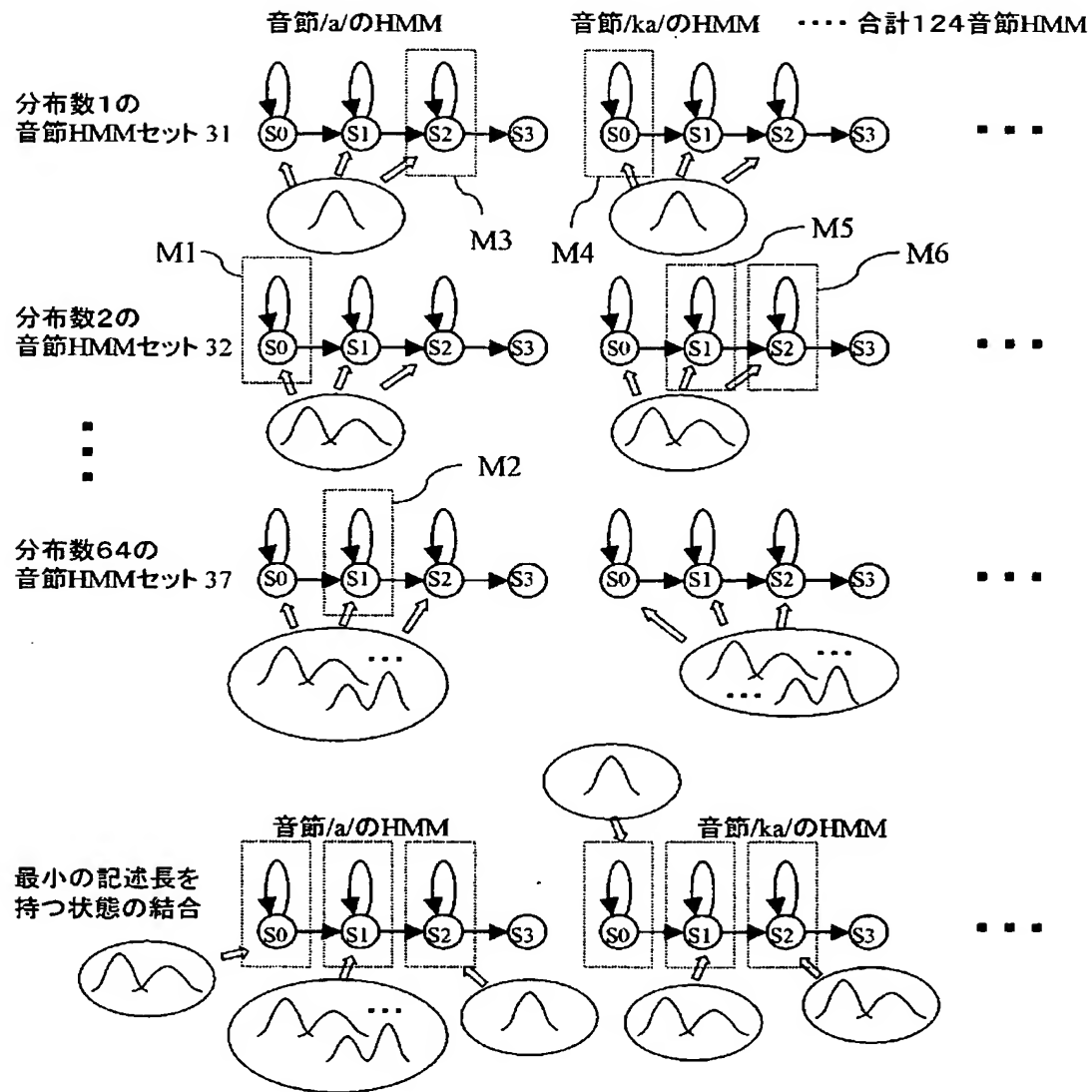
音節/a/のHMM (例)



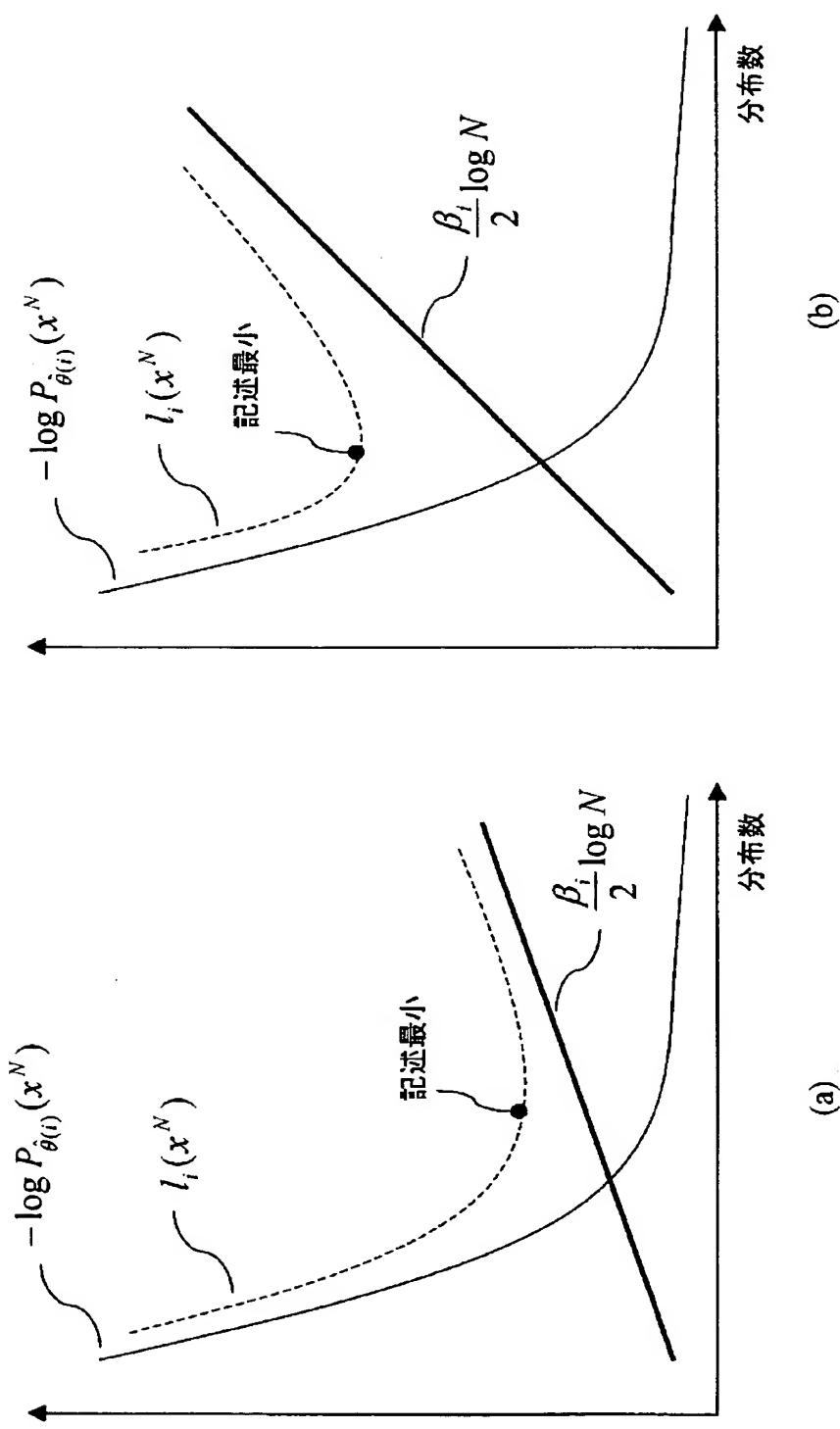
【図 7】



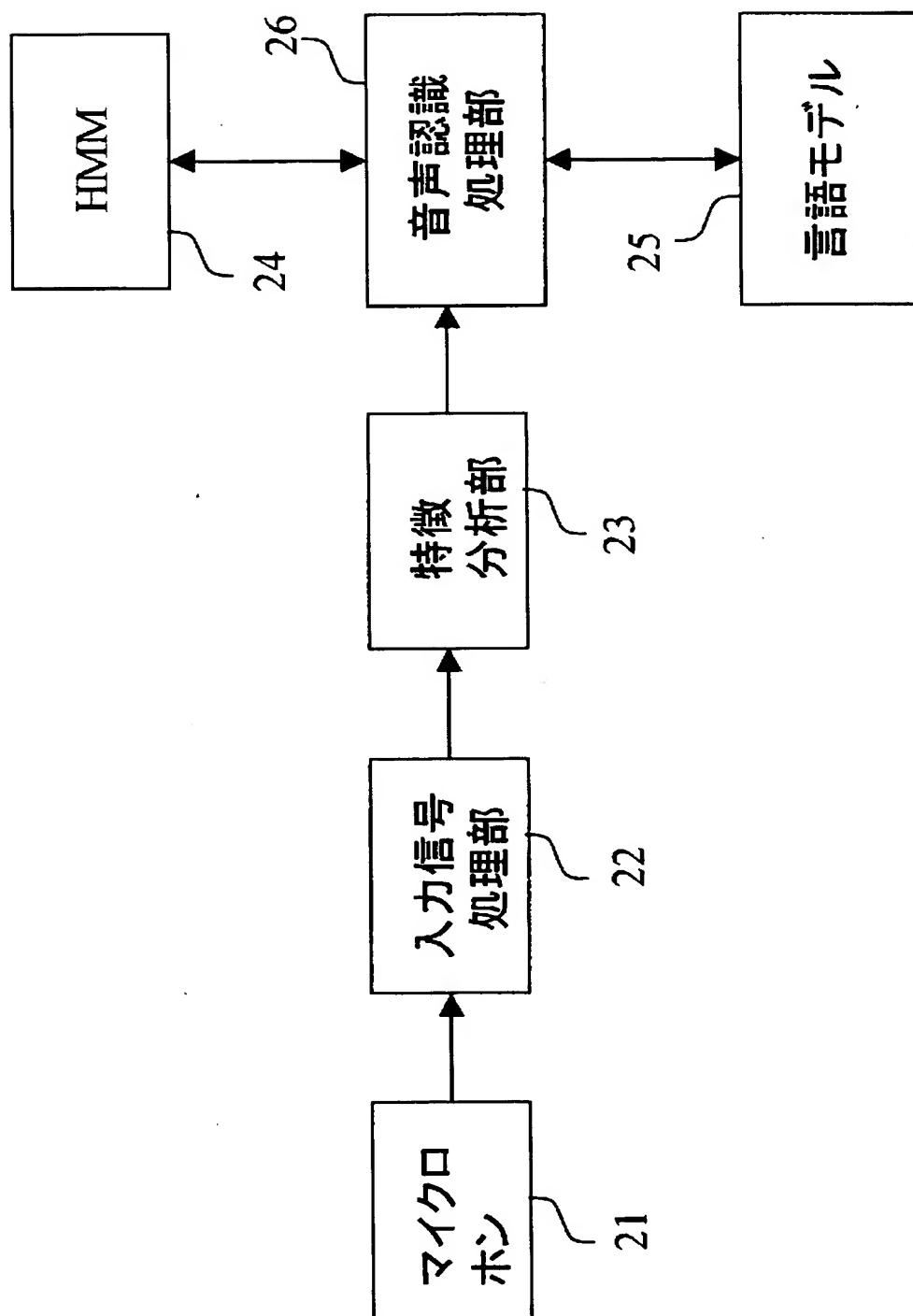
【図 8】



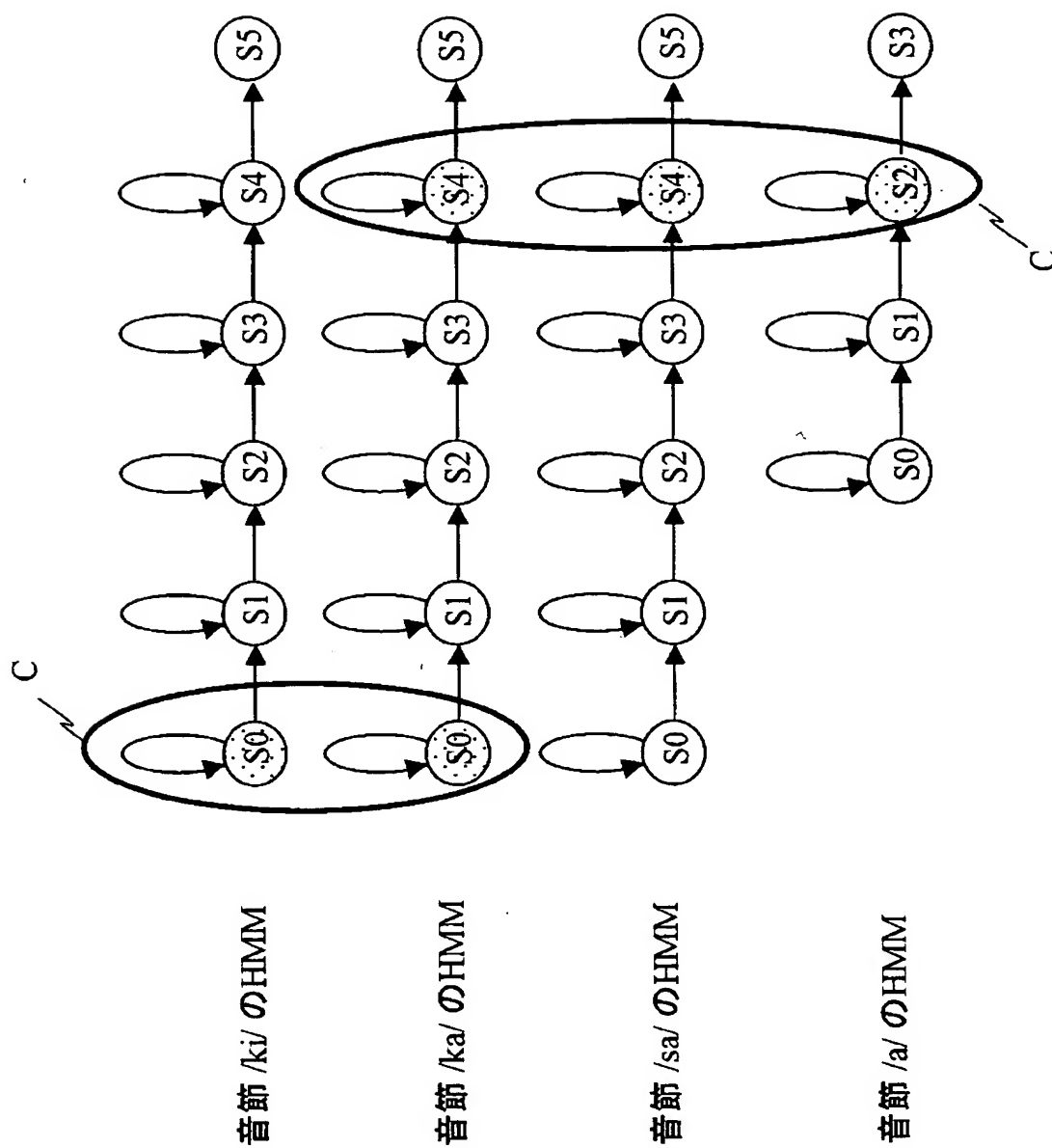
【図9】



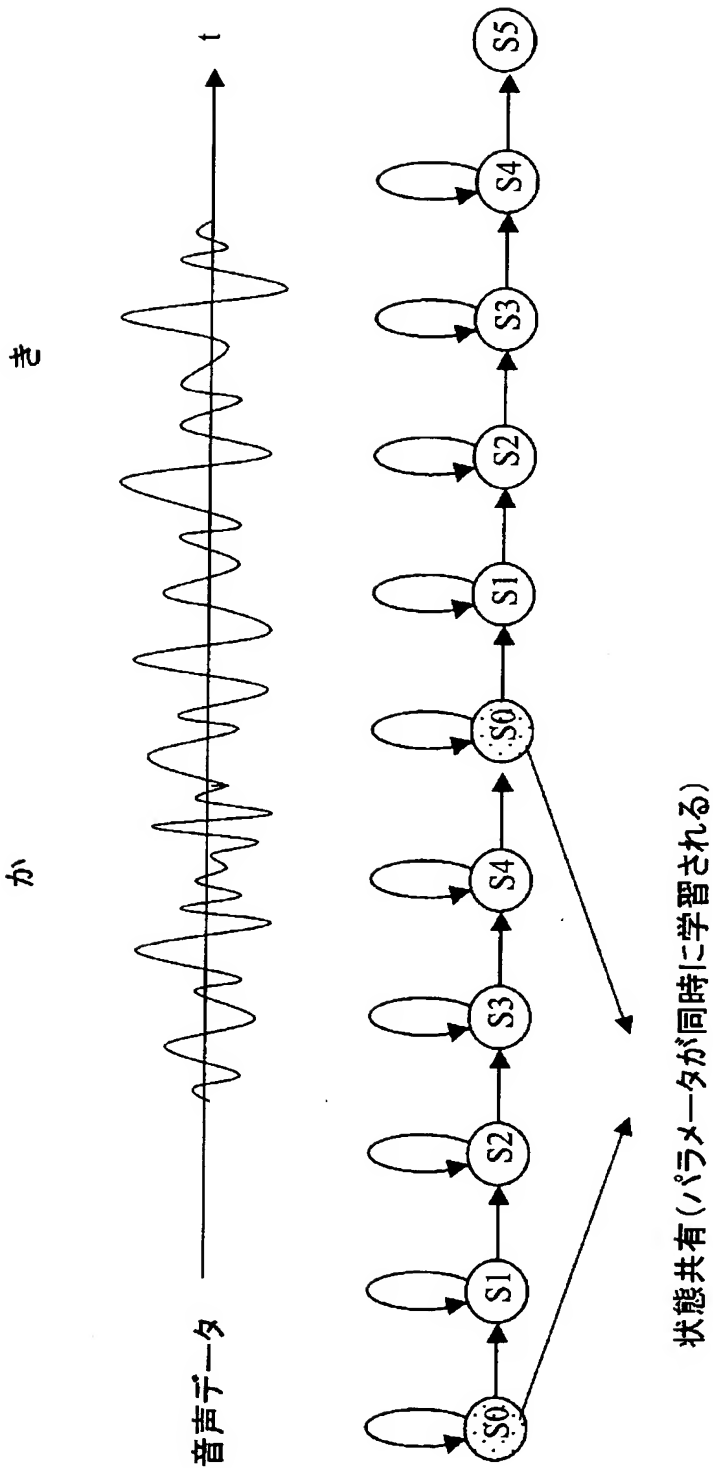
【図 10】



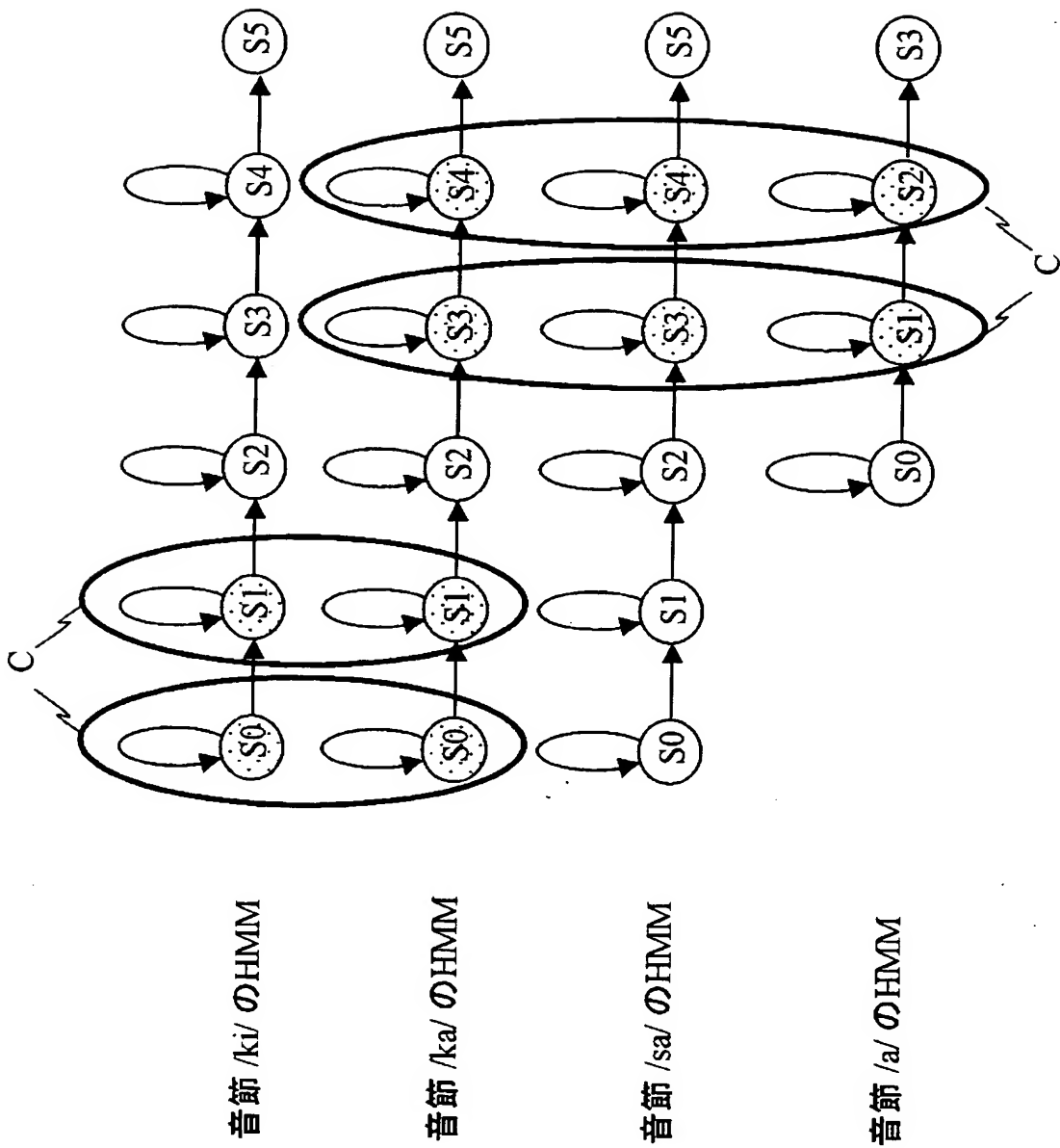
【図 11】



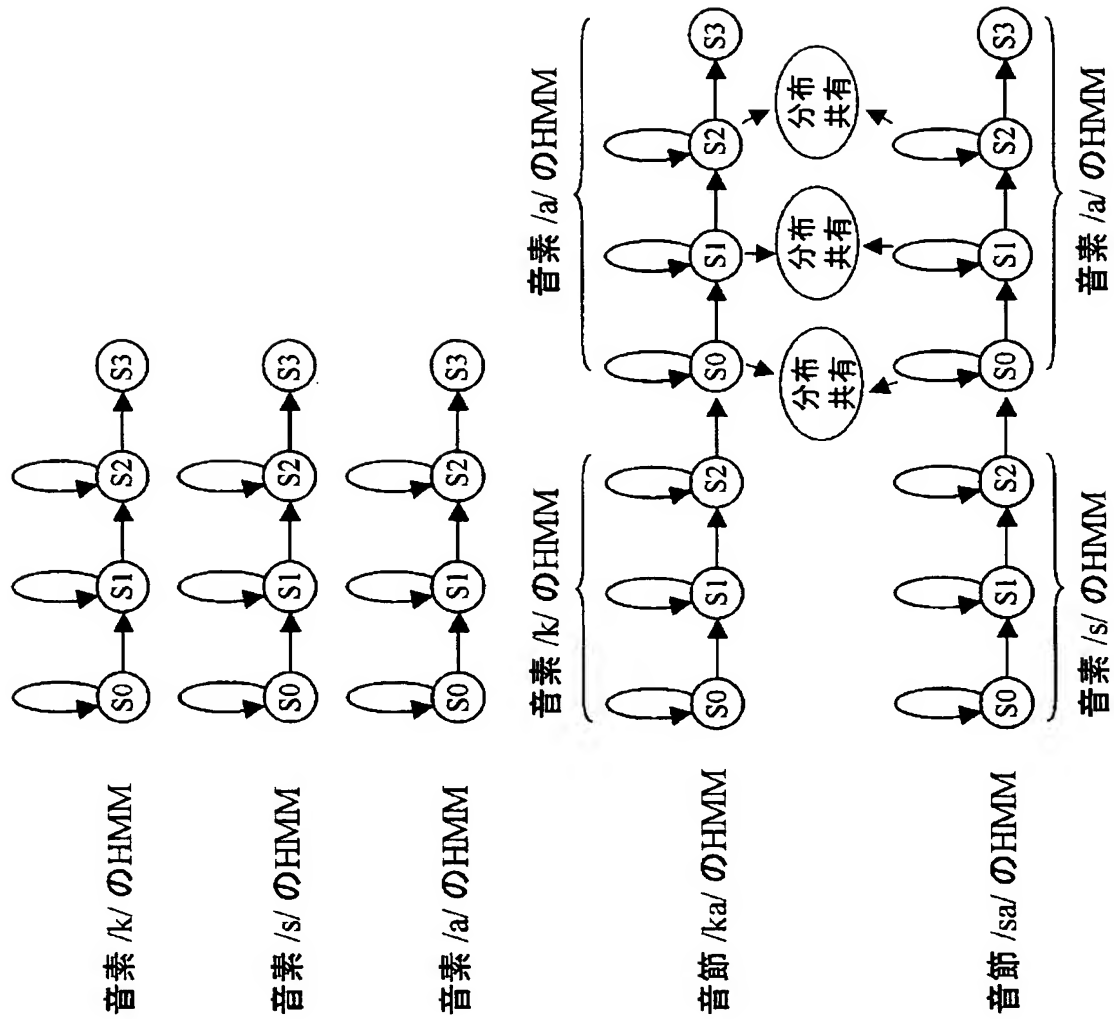
【図 12】



【図 13】



【図 14】



【書類名】 要約書

【要約】

【課題】 パラメータ数の削減を図り、しかも高い認識性能を得る HMM とするために HMM を構成する状態ごとにガウス分布数を最適化する。

【解決手段】 個々の音節 HMM を構成するそれぞれの状態のガウス分布数を 1 から最大分布数（分布数 6 4）とした HMM セットを学習用音声データ 1 を用いて学習し、学習された音節 HMM セット 3 1 ～ 3 7 のうち最大分布数に設定された音節 HMM セット 3 7 を用いて、個々の HMM の各状態とその HMM に対応する学習用音声データとをそれぞれの状態ごとにビタビアライメントをとる。そして、記述長計算部 6 がアライメントデータ 5 を用いて個々の HMM の各状態ごとに記述長を求め、記述長が最小となる分布数を持つ状態を状態選択部 8 によって選択する。そして、記述長が最小となる分布数を持つ状態によって個々の HMM を構築し、HMM 再学習部 9 によって、それらの HMM に対して学習用音声データを用いて再学習する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2 0 0 2 - 2 7 3 0 7 1
受付番号	5 0 2 0 1 4 0 2 9 8 3
書類名	特許願
担当官	第八担当上席 0 0 9 7
作成日	平成 1 4 年 9 月 2 0 日

< 認定情報・付加情報 >

【提出日】	平成14年 9月19日
-------	-------------

次頁無

特願 2 0 0 2 - 2 7 3 0 7 1

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 2 3 6 9]

1 . 変 更 年 月 日

1 9 9 0 年 8 月 2 0 日

[変 更 理 由]

新 規 登 録

住 所

東 京 都 新 宿 区 西 新 宿 2 丁 目 4 番 1 号

氏 名

セ イ コ ー エ プ ソ ン 株 式 会 社